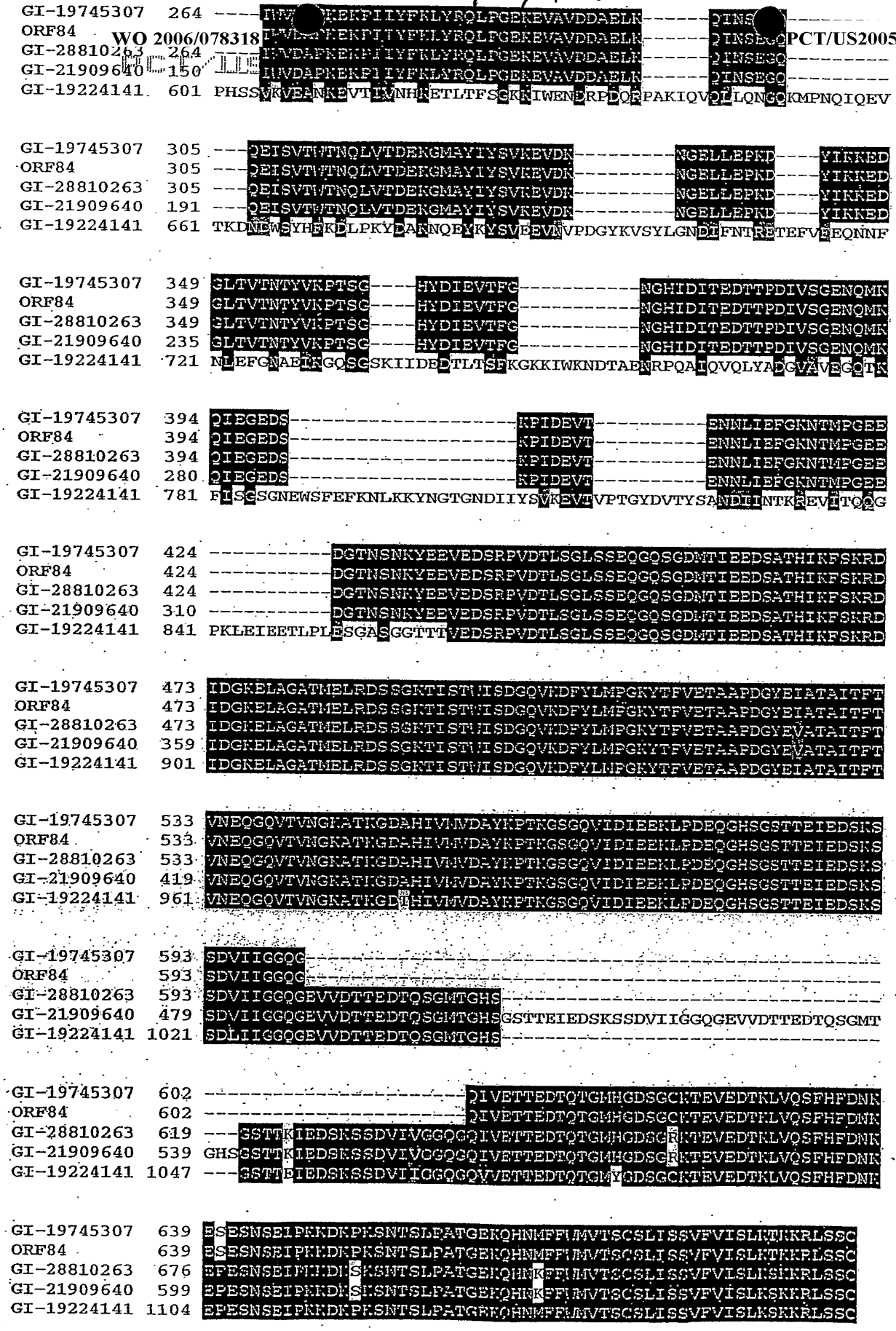

GI-19745307 264
ORF84
GI-28810263 264
GI-21909640 150
GI-19224141 601
GI-19745307 305
ORF84 305
GI-28810263 305
GI-21909640 191
GI-19224141 661
GI-19745307 349
ORF84 349
GI-28810263 349
GI-21909640 235
GI-19224141 721
GI-19745307 394
ORF84 394
GI-28810263 394
GI-21909640 280
GI-19224141 781
GI-19745307 424
ORF84 424
GI-28810263 424
GI-21909640 310
GI-19224141 841
GI-19745307 473
ORF84 473
GI-28810263 473
GI-21909640 359
GI-19224141 901
GI-19745307 533
ORF84 533
GI-28810263 533
GI-21909640 419
GI-19224141 961
GI-19745307 593
ORF84 593
GI-28810263 593
GI-21909640 479
GI-19224141 1021
GI-19745307 602
ORF84 602
GI-28810263 619
GI-21909640 539
GI-19224141 1047
GI-19745307 639
ORF84 639
GI-28810263 676
GI-21909640 599
GI-19224141 1104

FIGURE 54A

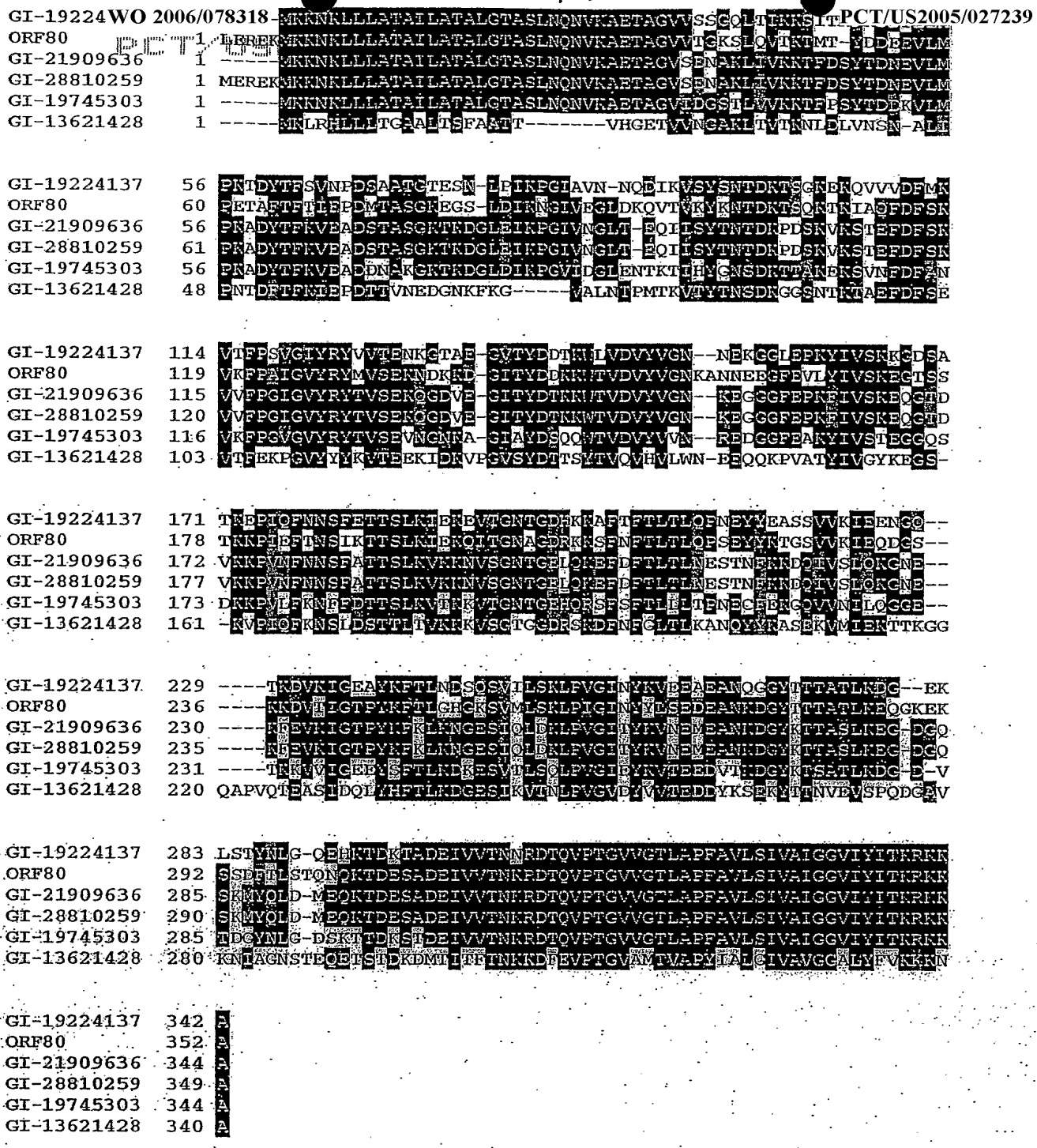

GI-19224137
ORF80
GI-21909636
GI-28810259
GI-19745303
GI-13621428

FIGURE 55

```
GI-21909638    1 ----------------------------------------MILTMLAFNQTVLAKDSTV
GI-28810261    1 ----------------------------------------MLFSVVMILTMLAFNQTVLAKDSTV
GI-19224139    1 ----------------------------------------MLFSVVMILTMLAFNQTVLAKDSTV
ORF82          1 LLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILTMLAFNQTVLAKDSTV
GI-19745305    1 ------------------------------MRKYWKMLFSVVMILTMLAFNQTVLAKDSTV

GI-21909638   20 QTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQYTYRV
GI-28810261   26 QTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQYTYRV
GI-19224139   26 QTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQYTYRV
ORF82         61 QTSISVENVLERAGDSTPFSVALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQYTYRV
GI-19745305   32 QTSISVENVLERAGDSTSFSVALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQYTYRV

GI-21909638   80 YQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPNWLVKPIP
GI-28810261   86 YQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPNWLVKPIP
GI-19224139   86 YQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
ORF82        121 YQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-19745305   92 YQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP

GI-21909638  140 PRQPNIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-28810261  146 PRQPNIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-19224139  146 PRQPNIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSKL
ORF82        181 PRQPNIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-19745305  152 PRQPDIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
```

FIGURE 56

```
!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913505.pep  from: 1 to: 1036  September 15, 2004 18:46

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M
GAS10394]

 TO: *.pep  Sequences:          56  Symbols:        22,803  Word Size: 2

 Scoring matrix: GenRunData:blosum50.cmp
 Variable pamfactor used
 Gap creation penalty: 12  Gap extension penalty: 2


Histogram Key:
 Each histogram symbol represents 1 search set sequences
 z-scores computed from opt scores

z-score obs     exp
        (=)     (*)

< 20      0      0:
  22      0      0:
  24      0      0:
  26      0      0:
  28      0      0:
  30      0      0:
  32      0      0:
  34      0      1:*
  36      0      2:  *
  38      0      3:   *
  40      0      4:    *
  42      0      5:     *
  44      0      5:     *
  46      5      5:====*
  48     10      5:====*=====
  50     12      4:===*========
  52      4      4:===*
  54      6      3:==*===
  56      4      3:==*=
  58      5      2:=*===
  60      3      2:=*=
  62      0      1:*
  64      1      1:*
  66      1      1:*
  68      0      1:*
  70      1      1:*
  72      0      0:
  74      0      0:
  76      1      0:=
  78      0      0:
  80      0      0:
  82      0      0:
  84      0      0:
  86      0      0:
  88      0      0:
  90      0      0:
  92      0      0:
  94      0      0:
  96      1      0:=
  98      0      0:
```

FIGURE 57

PCT/US05/27239

```
 100       0       0:
 102       0       0:
 104       0       0:
 106       1       0:=
 108       0       0:
 110       0       0:
 112       0       0:
 114       0       0:
 116       0       0:
 118       0       0:
>120       1       0:=

Joining threshold: 39, opt. threshold: 27, opt. width:  16, reg.-scaled


The best scores are:                      initl initn   opt    z-sc E(54)..

/home/morama/gas/pili/align/gi-50913505.pep    Begin: 1  End: 1036
! gi|50913505|ref|YP_059477.1| Collag...  6697  6697  6697  3452.1  9.6e-189
/home/morama/gas/pili/align/gi-19224141.pep    Begin: 48  End: 144
! gi|19224141|gb|AAL86412.1|AF447492_...    63   100   159   105.9   0.023
/home/morama/gas/pili/align/gi-21909640.pep    Begin: 147  End: 449
! gi|21909640|ref|NP_663908.1| protei...    35    35   136    96.3    0.08
/home/morama/gas/pili/align/gi-13621428.pep    Begin: 57  End: 318
! gi|13621428|gb|AAK33238.1| hypothet...    33    33    91    75.6     1.1
/home/morama/gas/pili/align/gi-50913506.pep    Begin: 33  End: 428
! gi|50913506|ref|YP_059478.1| Fimbri...    70   149    86    71.3     1.9
/home/morama/gas/pili/align/gi-13621432.pep    Begin: 14  End: 56
! gi|13621432|gb|AAK33241.1| conserve...    40    65    78    68.0     2.9
/home/morama/gas/pili/align/gi-19745301.pep    Begin: 241  End: 466
! gi|19745301|ref|NP_606437.1| putati...    52    52    73    64.8     4.3
/home/morama/gas/pili/align/gas15.pep    Begin: 492  End: 739
! GAS15 GAS15                              43    68    69    61.4     6.6
/home/morama/gas/pili/align/gi-21909636.pep    Begin: 176  End: 298
! gi|21909636|ref|NP_663904.1| conser...    31    31    62    60.8     7.1
/home/morama/gas/pili/align/gi-28810259.pep    Begin: 181  End: 303
! gi|28810259|dbj|BAC63197.1| hypothe...    31    31    62    60.7     7.2
/home/morama/gas/pili/align/gi-19224139.pep    Begin: 90  End: 143
! gi|19224139|gb|AAL86410.1|AF447492_...    43    43    54    58.9     8.9
/home/morama/gas/pili/align/gi-19745305.pep    Begin: 96  End: 149
! gi|19745305|ref|NP_606441.1| hypoth...    43    43    54    58.8       9
/home/morama/gas/pili/align/orf82.pep    Begin: 125  End: 178
! TRANSLATE of: orf82.seq check: 4296...    43    43    54    58.2     9.6
/home/morama/gas/pili/align/gi-21909638.pep    Begin: 84  End: 137
! gi|21909638|ref|NP_663906.1| hypoth...    43    43    52    58.0     9.9
\\End of List


gi-50913505.pep
/home/morama/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M
GAS10394]


SCORES   Init1: 6697  Initn: 6697  Opt: 6697  z-score: 3452.1 E(): 9.6e-189
>>/home/morama/gas/pili/align/gi-50913505.pep             (1036 aa)
 initn: 6697 init1: 6697 opt: 6697 Z-score: 3452.1 expect(): 9.6e-189
Smith-Waterman score: 6697;   100.0% identity in 1036 aa overlap
 (1-1036:1-1036)

                    10        20        30        40        50        60
gi-50913505. MYSRLKRELVIVINRKKKYKLIRLMVTVGLIFSQLVLPIRRLGLQMISTQTKVIPQEIVT
```

FIGURE 57A

```
gi-50913505. MYSRLKRELVIVINRKKKYKLIRLMVTVGLIFSQLVLPIRRLGLQMISTQTKVIPQEIVT
                    10        20        30        40        50        60

                    70        80        90       100       110       120
gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAI
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAI
                    70        80        90       100       110       120

                   130       140       150       160       170       180
gi-50913505. VTMALELRKQGLSIVDTKIVRIQSSTNQRNDITTTLTFKNGLSLEGASTEANDPNVRVGI
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. VTMALELRKQGLSIVDTKIVRIQSSTNQRNDITTTLTFKNGLSLEGASTEANDPNVRVGI
                   130       140       150       160       170       180

                   190       200       210       220       230       240
gi-50913505. VNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIVSTTRLKEEQTISLDSYGELVID
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. VNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIVSTTRLKEEQTISLDSYGELVID
                   190       200       210       220       230       240

                   250       260       270       280       290       300
gi-50913505. GAVGLSQKDRPPYSKPITVNILKPKLSSIESSLDSKDFEIVKTIDNLYTWDDQFYLLDFI
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. GAVGLSQKDRPPYSKPITVNILKPKLSSIESSLDSKDFEIVKTIDNLYTWDDQFYLLDFI
                   250       260       270       280       290       300

                   310       320       330       340       350       360
gi-50913505. SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIG
                   310       320       330       340       350       360

                   370       380       390       400       410       420
gi-50913505. AAPALIQPRSFRSLTPRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDE
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. AAPALIQPRSFRSLTPRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDE
                   370       380       390       400       410       420

                   430       440       450       460       470       480
gi-50913505. HDTSDLYRLYLDMTGKKNPLDILVVVDKSGSMQEGIGSVQRYRYYAQRWDDYYSQWVYHG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. HDTSDLYRLYLDMTGKKNPLDILVVVDKSGSMQEGIGSVQRYRYYAQRWDDYYSQWVYHG
                   430       440       450       460       470       480

                   490       500       510       520       530       540
gi-50913505. TFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNGLLQRFVNINPENKL
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. TFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNGLLQRFVNINPENKL
                   490       500       510       520       530       540

                   550       560       570       580       590       600
gi-50913505. SVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. SVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
                   550       560       570       580       590       600

                   610       620       630       640       650       660
gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRS
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRS
                   610       620       630       640       650       660
```

FIGURE 57B

```
                670       680       690       700       710       720
gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELE
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELE
                670       680       690       700       710       720

                730       740       750       760       770       780
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
                730       740       750       760       770       780

                790       800       810       820       830       840
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAYEKYKDNEGRYS
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAYEKYKDNEGRYS
                790       800       810       820       830       840

                850       860       870       880       890       900
gi-50913505. EMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. EMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
                850       860       870       880       890       900

                910       920       930       940       950       960
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYYLYETKAKLGYTLP
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYYLYETKAKLGYTLP
                910       920       930       940       950       960

                970       980       990      1000      1010      1020
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTA
                970       980       990      1000      1010      1020

               1030
gi-50913505. TVALLFYRRQHRKKQY
             ||||||||||||||||
gi-50913505. TVALLFYRRQHRKKQY
               1030


gi-50913505.pep
/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]


SCORES  Init1: 63   Initn: 100   Opt: 159   z-score: 105.9 E(): 0.023
>>/home/morama/gas/pili/align/gi-19224141.pep              (1161 aa)
 initn: 100 init1:  63 opt: 159 Z-score: 105.9 expect(): 0.023
Smith-Waterman score: 159;    36.7% identity in 98 aa overlap
 (895-990:48-144)

              870       880       890       900       910       920
gi-50913505. SDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVDADNNQKKLAGVEFELRKEDKK-IV
                                        :|    :|:|:| | |: | | ::| |
gi-19224141. FILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKGTS
              20        30        40        50        60        70

               930       940       950       960       970       980
```

FIGURE 57C

```
gi-50913505. WEKGTTGSNGQLNFKYLQKGKTYYLYETKAKLGYTLPENPWEVAVANNGDIK-VKHPIEG
              :  |::::| :: : || | || | |  |  ||      | |:| :||  | |::| :|
gi-19224141. VQTFTSNDKGIVDAQNLQPG-TYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNG
              80        90        100       110       120       130

                  990       1000      1010      1020      1030
gi-50913505. ELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTATVALLFYRRQHRKKQY
             |: || ||
gi-19224141. EIISKAGSKDVSSSLQLENPKMSVVSKYGKTEVSSGAADFYRNHAAYFKMSFELKQKDKS
              140       150       160       170       180       190


gi-50913505.pep
/home/morama/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP_663908.1| protein F2 like fibronectin-binding protein [Strept
ococcus pyogenes MGAS315]


SCORES   Init1: 35    Initn: 35    Opt: 136   z-score: 96.3  E(): 0.08
>>/home/morama/gas/pili/align/gi-21909640.pep              (656 aa)
 initn:  35 init1:  35 opt: 136 Z-score: 96.3 expect(): 0.08
Smith-Waterman score: 148;    24.5% identity in 339 aa overlap
 (686-1005:147-449)

             660       670       680       690       700             710
gi-50913505. NVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVL----KYLSGEEHYY
                                              :| |  |: : :|::     : | ||::
gi-21909640. TNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEV-
              120       130       140       150       160       170

                 720       730       740       750       760       770
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQ
              :| |||::      :::   :::::: : :| |   |::  : :   ::|: : |:|
gi-21909640. -AVDDAELKQ-----INSEGQQEISVTWT-NQLVT--DEKGMAYIYSVKEVDKNGELLEP
               180             190       200       210       220

              780       790       800       810       820       830
gi-50913505. KDQVQ-EAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAYE
             || :: | |  : :  |   | ||  :  : :|| : : :|   |   : :: ::::
gi-21909640. KDYIKKEDGLTVTNTYV---KPTSG-HYDIEVTFGNGH-ID--ITEDTTPDIVSGEN---
              230       240          250       260          270

                840       850       860       870
gi-50913505. KYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDA--SVNY--MADGR----------
             :|: ||: |:  |  |:  :|   ||: :|:: |:  | :|  : |:|
gi-21909640. QMKQIEGEDSKPIDEVTE--NNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSS
              280       290       300       310       320       330

             880       890       900       910       920       930
gi-50913505. EQKLPYKHPVIQVKTVPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLN
             ||      : : ::: | |:| [ |:  |:|||: :|||  : | :   :|  |:||::
gi-21909640. EQGQSGDMTIEEDSATHIKFSKRDIDG--KELAGATMELRDSSGKTI---STWISDGQVK
              340       350       360       370       380

             940       950       960       970       980       990
gi-50913505. FKYLQKGKTYYLYETKAKLGYTLPENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNY
              ||: || | : || |  ||        |||:| :  :: :  :  : |:  |:  |
gi-21909640. DFYLMPGK-YTFVETAAPDGY--------EVATAITFTVNEQGQVTVNGKATKGDAHIVMV
          390       400               410       420       430       440

            1000      1010      1020      1030
gi-50913505. KIYQLPSSGGRGSQIFIIVGSMTATVALLFYRRQHRKKQY
```

FIGURE 57D

```
                  |: |::|:
gi-21909640. DAYK-PTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGM
                    450       460       470       480       490       500


gi-50913505.pep
/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]


SCORES   Init1: 33    Initn: 33    Opt: 91    z-score: 75.6  E(): 1.1
>>/home/morama/gas/pili/align/gi-13621428.pep             (340 aa)
 initn:  33 init1:  33 opt:  91 Z-score: 75.6 expect():  1.1
Smith-Waterman score: 95;    19.9% identity in 271 aa overlap
 (568-819:57-318)

               540       550       560       570       580       590
gi-50913505. NKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTAL
                                                :|: : ::|:: :|| : |: :   | |
gi-13621428. VNGAKLTVTKNLDLVNSNALIPNTDFTFKIEPDTTVNEDGNKFKGVALNTPMTKVTYTNS
              30        40        50        60        70        80

               600       610       620       630       640              650
gi-50913505. HNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYR-------SGNGS
              ::|:| ::| :  :|:  |       |:    | ||   :  :|        : : :
gi-13621428. DKGGSNTKTAEFDFSEVTFEKPGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQ
               90        100       110       120       130       140

                    660       670       680       690       700
gi-50913505. SNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTAS-SSPVVLKYLSGEEH
             :   : ::  :|||| : :||   :|:  :| |:| ::: :: |:   :      ::
gi-13621428. KPVATYIVGYKEGSKVPI-QFKN---SLDSTTLTVKKKVSGTGGDRSKDFNFGLTLKANQ
              150       160           170       180       190       200

           710       720           730           740       750
gi-50913505. YYGITDTAELEKTLN----KIVEDSKLSQL---GISDSLSQYVDYYDKQPDVLVT-----R
             ||  :: : :||| :     :  :::::||    ::|: |  |        | :||    :
gi-13621428. YYKASEKVMIEKTTKGGQAPVQTEASIDQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYK
                  210       220       230       240       250       260

             760       770       780       790       800       810
gi-50913505. KSKVNDETEILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTL
             : | : ::|:  |   |:: : :  :: : | |  :      :|:| |:|:::|    ::|:
gi-13621428. SEKYTTNVEVSPQDGAVKNIAGNSTEQETSTDKDMT-----ITFTNKKDFEVPTGVAMTV
                  270       280       290           300       310

          820       830       840       850       860       870
gi-50913505. SFNVKASDEAYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQ
             :
gi-13621428. APYIALGIVAVGGALYFVKKKNA
             320       330       340


gi-50913505.pep
/home/morama/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes
 MGAS10394]


SCORES   Init1: 70    Initn: 149   Opt: 86    z-score: 71.3  E(): 1.9
>>/home/morama/gas/pili/align/gi-50913506.pep             (556 aa)
```

FIGURE 57E

```
initn: 149 init1:  70 opt:  86 Z-score: 71.3 expect():  1.9
Smith-Waterman score: 120;    21.5% identity in 469 aa overlap
(503-966:33-428)

                480       490       500       510       520
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNG-----L
                                                 |:: || :: :||::: :: :::|     |
gi-50913506. NRRETVREKILITAKKLMLACLAILAVVGLGMTRVS-ALSKDDTAQLKITNIEGGPTVTL
                10        20        30        40        50        60

                530       540       550       560       570       580
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS
              :   ::   |  | |:|:    |  |    : :| :     :: |: ::  :| :||::
gi-50913506. YKIGEGVYNTNGDSFINFK----YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSTEN
                70        80              90        100       110

                590       600       610       620       630       640
gi-50913505. LLDPNTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSG
             :   |  :: :  |:: : |||  :          |||    |::: :    | || :  :
gi-50913506. VSISNGTATYNARGASVYIALLTGAT-------DGRTYNPILLAAS----YNGEGNLVTK
              120       130       140             150       160

                650       660       670       680       690       700
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGE
             | :|  ::|   :|  :::|    |:  |:::   |:  |:|: |:| : |:   |: |
gi-50913506. NIDS--KSNYLYGQ--TSVA----KSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFE
                170       180             190       200       210

                710       720       730       740       750       760
gi-50913505. EHYYGITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETE
                |         ::::|| |          :||::|: : :  :  :: |  |:|: : ||
gi-50913506. LPSY-------TKEAVNKTVY---------VSDNMSEGLTF--NFNSLTVEWKGKMANITE
             220               230               240       250       260

                770       780       790       800       810       820
gi-50913505. ILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDE
                        | |::.:|      :|:  |: ::  |:     ::| |:
gi-50913506. ----------------DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
                               270       280       290

                830       840       850       860       870       880
gi-50913505. AYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVI
             :|:   :|::  :| |: :      :::  ||:  :| | : :  ::|  :|   |
gi-50913506. SYKAVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDK-GNGITSKEDSK----
             300       310       320       330       340       350

                890       900       910       920       930       940
gi-50913505. QVKTVPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYY
              | |  |:| |||:  ::  | |: | :   ::|:: :  ||::||     :::|| |
gi-50913506. IVYTYQIAFRKVDS-VSKTPLIGAIFGVYDTSNKLI-DIVTTNKNGYAISTQVSSGK-YK
                  360       370       380       390       400

                950       960       970       980       990      1000
gi-50913505. LYETKAKLGYTLPENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGR
             : | ||  ||:|  : :|::
gi-50913506. IKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGI
           410       420       430       440       450       460


gi-50913505.pep
/home/morama/gas/pili/align/gi-13621432.pep

gi|13621432|gb|AAK33241.1| conserved hypothetical protein [Streptococcus pyogene
```

FIGURE 57F

s]

```
SCORES   Init1: 40    Initn: 65    Opt: 78    z-score: 68.0  E(): 2.9
>>/home/morama/gas/pili/align/gi-13621432.pep              (450 aa)
 initn:  65 init1:  40 opt:  78 Z-score: 68.0 expect():  2.9
Smith-Waterman score: 78;    37.0% identity in 46 aa overlap
 (368-411:14-56)

             340       350       360       370       380       390
gi-50913505. KGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRSLTPRSTRMK--RSAPVEKFEGELE
                                          |   :::| |  ::|  |:| : :|| ||
gi-13621432.                      MTRTNYQKKRMTCPVETEDITYRRKKIKGRRQAILAQFEPELV
                                         10        20        30        40

              400       410       420       430       440       450
gi-50913505. HHKRIDYLGDNQNNPDTTIDDKEDEHDTSDLYRLYLDMTGKKNPLDIEVVVDKSGSMQEG
             ||: |   ||: : ||
gi-13621432. HHELI---GDSCTCPDCHGTLTEIGSVVQRQELVFIPAQLKRINHVQHAYKCQTCSDNSL
                   50        60        70        80        90       100


gi-50913505.pep
/home/morama/gas/pili/align/gi-19745301.pep

gi|19745301|ref|NP_606437.1| putative collagen binding protein [Streptococcus py
ogenes MGAS8232]


SCORES   Init1: 52    Initn: 52    Opt: 73    z-score: 64.8  E(): 4.3
>>/home/morama/gas/pili/align/gi-19745301.pep              (524 aa)
 initn:  52 init1:  52 opt:  73 Z-score: 64.8 expect():  4.3
Smith-Waterman score: 95;    23.7% identity in 245 aa overlap
 (759-989:241-466)

            730       740       750       760       770       780
gi-50913505. DSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAG-KDIIDKVV
                                          ::::|||  :| |  :| | | :|  |:::
gi-19745301. ETIDPDFNEGKEIKYTHILGADLFSYANNPRASTNDE--LLSQVKKVLEKGYRD--DSTT
                   220       230       240       250       260

             790       800       810       820       830       840
gi-50913505. FTPKTTSQPKGKVTLT---FKSDYKVDD--EYTYTLSFNVKASDEAYEKYKDNEGRYSEM
             ::  |: : :: : |:  | :: :|:  :|   ::::| : : |     | | :
gi-19745301. YANLTSVEFRAATQLAIYYFTDSVDLDNLADYHGFGALTTEALNATKEIVAYAEDRANLP
              270       280       290       300       310       320

                   850       860       870       880       890
gi-50913505. GDSDTDY---GTNQTSS--GKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFT
             : |: |:   ::|: :|  |    | :| ::: | | :   :|  | :   |||  |:
gi-19745301. NISNLDFYVPNSNKYQSLIGTQYHP-ESLVDIIRMEDKQAPIIPITHKLTISKTVTGTI-
              330       340       350       360       370       380

             900       910       920       930       940       950
gi-50913505. KVDADNNQKKLAGVEFELRKEDKKIVWEKGTTGSN-GQLNFKYLQKGK-TYYLYETKAKL
               ||  :||  : |::|:: | : :  :|| :| |:|:   :  || |: | : :: :
gi-19745301. ---AD--KKKEFNFEIHLKSSDGQAI--SGTYPTNSGELT---VTDGKATFTLKDGESLI
                     390       400       410       420       430

              960       970       980       990      1000      1010
gi-50913505. GYTLPEN-PWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFII
               || :  :|:: :: :| :|:  ::|: :: ||:
gi-19745301. VEGLPSGYSYEITETGASDYEVS--VNGK-NAPDGKATKASVKEDETITFENRKDLVPPT
```

FIGURE 57G

PCT/US05/27239

```
                440       450       460       470       480       490

              1020      1030
gi-50913505.  VGSMTATVALLFYRRQHRKKQY

gi-19745301.  GLTTDGAIYLWLLLLVLLGLWVWLIGRKGLKND
                    500       510       520


gi-50913505.pep
/home/morama/gas/pili/align/gas15.pep

GAS15 GAS15


SCORES   Init1: 43    Initn: 68    Opt: 69    z-score: 61.4  E(): 6.6
>>/home/morama/gas/pili/align/gas15.pep                    (762 aa)
 initn:  68 init1:  43 opt:  69 Z-score: 61.4 expect():  6.6
Smith-Waterman score: 100;    21.4% identity in 252 aa overlap
 (641-873:492-739)

                   620       630       640       650       660
gi-50913505.  AKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQ--EGSKLAI
                                                 | |||  :|:: : :::|::|   :::|||
gas15.pep     HIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYRE-KGQAIEYSGLTETQLRAATQLAI
                   470       480       490       500       510       520

              670       680       690       700       710       720
gi-50913505.  DEF--KARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELEKTLNKI
                | :|:  : :: :     |:|::| : : :::| : : :  :|| :  : ||
gas15.pep     YYFTDSAELDKDKLKDYHGFGDMNDSTLAVAKILVEY-AQDSNPPQLTDLDFFIPNNNKY
                    530       540       550       560       570

               730       740       750             760           770
gi-50913505.  VEDSKLSQLGISDSLSQYVDYYDKQPDVLVT-----RKSKVN---DETEILYQKDQVQEA
                :| ::     ::| : : : ||:  : ||     ||: ::   |:|: :: : ::::
gas15.pep     --QSLIGTQWHPEDLVDIIRMEDKKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNN
              580       590       600       610       620       630

             780       790       800       810       820       830
gi-50913505.  GKDIIDKVVFTPKTTSQPK-GKVTLTFKSDYKVDDE-YTYTLSFNVKASDEAYEKYKDNE
              :::::::| | ||: : | ||:|:::|   ::  :      |: || :|    | | |
gas15.pep     KQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQGLPEGYSYLVKETDSEGYKVKVNS
              640       650       660       670       680       690

               840       850             860       870       880       890
gi-50913505.  GRYSEMGDSDTDYGTNQT----SSGKGGLPSNSDASVN-YMADGREQKLPYKHPVIQVKT
                : ::   | |  :::|    :: :  :|:: | ::| |:|
gas15.pep     QEVANATVSKTGITSDETLAFENNKEPVVPTGVDQKINGYLALIVIAGISLGIWGIHTIR
              700       710       720       730       740       750

                   900       910       920       930       940       950
gi-50913505.  VPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYYLYET

gas15.pep     IRKHD
              760


gi-50913505.pep
/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyoge
nes MGAS315]
```

FIGURE 57H

```
SCORES   Init1: 31    Initn: 31    Opt: 62    z-score: 60.8  E(): 7.1
>>/home/morama/gas/pili/align/gi-21909636.pep              (344 aa)
 initn:  31 init1:  31 opt:  62 Z-score: 60.8 expect():  7.1
Smith-Waterman score: 71;    22.9% identity in 131 aa overlap
 (181-305:176-298)

                    160       170       180       190       200       210
gi-50913505.  DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGR
                                                 || |::  | :  :|::::|:      ||:
gi-21909636.  VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGN------TGE
                150       160       170       180       190

                         220       230       240       250       260
gi-50913505.  LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKPKL
              | |:    : :  :| :|::| :||:: |:  ::  :|   | :   :: | :: |   :
gi-21909636.  LQKEFDFTLTLNESTNFKKDQIVSLQK-GNEKFEVKIGTPYKFKLKNGESIQLDKLPVGI
           200       210       220       230       240       250

              270       280       290       300       310       320
gi-50913505.  SSIESSLDSKDFEIVKTIDNLYTWDDQ--FYLLDFISKQYEVLKTDYQSAKDSTPQTRDI
              :   : ::::  :  ||  :|   | |  :| ||: :|  |
gi-21909636.  TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESADEIVVTNKRDTQVPTGV
             260       270       280       290       300       310

                330       340       350       360       370       380
gi-50913505.  LFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRSLTPRSTRMKRS

gi-21909636.  VGTLAPFAVLSIVAIGGVIYITKRKKA
              320       330       340


gi-50913505.pep
/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]


SCORES   Init1: 31    Initn: 31    Opt: 62    z-score: 60.7  E(): 7.2
>>/home/morama/gas/pili/align/gi-28810259.pep              (349 aa)
 initn:  31 init1:  31 opt:  62 Z-score: 60.7 expect():  7.2
Smith-Waterman score: 71;    22.9% identity in 131 aa overlap
 (181-305:181-303)

                    160       170       180       190       200       210
gi-50913505.  DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGR
                                                 || |::  | :  :|::::|:      ||:
gi-28810259.  VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGN------TGE
                160       170       180       190       200

                         220       230       240       250       260
gi-50913505.  LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKPKL
              | |:    : :  :| :|::| :||:: |:  ::  :|   | :   :: | :: |   :
gi-28810259.  LQKEFDFTLTLNESTNFKKDQIVSLQK-GNEKFEVKIGTPYKFKLKNGESIQLDKLPVGI
              210       220       230       240       250       260

              270       280       290       300       310       320
gi-50913505.  SSIESSLDSKDFEIVKTIDNLYTWDDQ--FYLLDFISKQYEVLKTDYQSAKDSTPQTRDI
              :   : ::::  :  ||  :|   | |  :| ||: :|  |
gi-28810259.  TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESADEIVVTNKRDTQVPTGV
              270       280       290       300       310       320
```

FIGURE 57I

PCT/US05/27239

```
                330       340       350       360       370       380
gi-50913505. LFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRSLTPRSTRMKRS

gi-28810259. VGTLAPFAVLSIVAIGGVIYITKRKKA
                  330       340


gi-50913505.pep
/home/morama/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]


SCORES   Init1: 43    Initn: 43    Opt: 54    z-score: 58.9  E(): 8.9
>>/home/morama/gas/pili/align/gi-19224139.pep              (189 aa)
 initn:  43 init1:  43 opt:  54 Z-score: 58.9 expect():  8.9
Smith-Waterman score: 54;    31.6% identity in 57 aa overlap
 (742-796:90-143)

                   720       730       740       750       760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL
                                          ||  ||       ||||     : ||   |
gi-19224139. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
             60        70        80        90       100       110

           770       780       790       800       810       820
gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY
               |   ::|| :  : ::| ||   :|
gi-19224139. VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPLPLAGEVKSLLGILSIVLLGL
              120       130       140       150       160       170


gi-50913505.pep
/home/morama/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP_606441.1| hypothetical protein [Streptococcus pyogenes MGAS82
32]


SCORES   Init1: 43    Initn: 43    Opt: 54    z-score: 58.8  E(): 9
>>/home/morama/gas/pili/align/gi-19745305.pep              (195 aa)
 initn:  43 init1:  43 opt:  54 Z-score: 58.8 expect():    9
Smith-Waterman score: 54;    31.6% identity in 57 aa overlap
 (742-796:96-149)

                   720       730       740       750       760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL
                                          ||  ||       ||||     : ||   |
gi-19745305. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
                 70        80        90       100       110       120

           770       780       790       800       810       820
gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY
               |   ::|| :  : ::| ||   :|
gi-19745305. VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPDIPKTPLPLAGEVKSLLGILSIVLLGL
                 130       140       150       160       170       180


gi-50913505.pep
/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
generated symbols 1 to: 224.
```

FIGURE 57J

PCT/US05/27239

```
GETSEQ from morama, September 13, 2004 17:09.


SCORES   Init1: 43    Initn: 43    Opt: 54    z-score: 58.2  E(): 9.6
>>/home/morama/gas/pili/align/orf82.pep                    (224 aa)
 initn:  43 init1:  43 opt:  54 Z-score: 58.2 expect():  9.6
Smith-Waterman score: 54;    31.6% identity in 57 aa overlap
 (742-796:125-178)

                720       730       740       750       760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL
                                          ||  ||       ||||    : ||   |
orf82.pep    ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
                100       110       120       130       140          150

           770       780       790       800       810       820
gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY
               |   ::|| :  : ::| ||   :|
orf82.pep    VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPLPLAGEVKSLLGILSIVLLGL
                     160       170       180       190       200       210


gi-50913505.pep
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS31
5]


SCORES   Init1: 43    Initn: 43    Opt: 52    z-score: 58.0  E(): 9.9
>>/home/morama/gas/pili/align/gi-21909638.pep              (183 aa)
 initn:  43 init1:  43 opt:  52 Z-score: 58.0 expect():  9.9
Smith-Waterman score: 52;    31.6% identity in 57 aa overlap
 (742-796:84-137)

                720       730       740       750       760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL
                                          ||  ||       ||||    : ||   |
gi-21909638. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
                    60        70        80        90       100          110

           770       780       790       800       810       820
gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY
               |   ::|| :  : ::| ||   :|
gi-21909638. VAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNIPKTPLPLAGEVKSLLGILSIVLLGL
                    120       130       140       150       160       170


! Distributed over 1 thread.
!      Start time: Wed Sep 15 18:46:18 2004
! Completion time: Wed Sep 15 18:46:21 2004

! CPU time used:
!        Database scan:  0:00:00.1
! Post-scan processing:  0:00:02.2
!       Total CPU time:  0:00:02.3
! Output File: gi-50913505.fasta
```

FIGURE 57K

PCT/US05/27239

```
!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913506.pep  from: 1 to: 556  September 15, 2004 18:45

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes
 MGAS10394]

 TO: *.pep  Sequences:          56  Symbols:         22,803  Word Size: 2

 Scoring matrix: GenRunData:blosum50.cmp
 Variable pamfactor used
 Gap creation penalty: 12  Gap extension penalty: 2


Histogram Key:
 Each histogram symbol represents 1 search set sequences
 z-scores computed from opt scores

z-score obs    exp
        (=)    (*)

< 20      0      0:
  22      0      0:
  24      0      0:
  26      0      0:
  28      0      0:
  30      0      0:
  32      0      0:
  34      0      1:*
  36      0      2: *
  38      0      3:  *
  40      0      4:   *
  42      0      5:    *
  44      0      5:    *
  46      1      5:=   *
  48      8      5:====*===
  50     24      5:====*==================
  52      2      4:== *
  54      3      3:==*
  56      1      3:= *
  58      2      2:=*
  60      3      2:=*=
  62      0      2: *
  64      2      1:*=
  66      2      1:*=
  68      1      1:*
  70      2      1:*=
  72      0      0:
  74      3      0:===
  76      1      0:=
  78      0      0:
  80      0      0:
  82      0      0:
  84      0      0:
  86      0      0:
  88      0      0:
  90      0      0:
  92      0      0:
  94      0      0:
  96      0      0:
  98      0      0:
```

FIGURE 58

```
 100      0      0:
 102      0      0:
 104      0      0:
 106      0      0:
 108      0      0:
 110      0      0:
 112      0      0:
 114      0      0:
 116      0      0:
 118      0      0:
>120      1      0:=

Joining threshold: 37, opt. threshold: 25, opt. width:  16, reg.-scaled


The best scores are:                      init1 initn   opt    z-sc E(55)..

/home/morama/gas/pili/align/gi-50913506.pep    Begin: 1  End: 556
! gi|50913506|ref|YP_059478.1| Fimbri... 3454  3454  3454  1016.2  4.7e-53
/home/morama/gas/pili/align/orf84.pep    Begin: 316  End: 567
! TRANSLATE of: orf84.seq check: 7868...   57    83   135    75.1     1.2
/home/morama/gas/pili/align/gi-19745307.pep    Begin: 316  End: 567
! gi|19745307|ref|NP_606443.1| protei...   57    83   135    75.1     1.2
/home/morama/gas/pili/align/gi-21909640.pep    Begin: 202  End: 524
! gi|21909640|ref|NP_663908.1| protei...   56    81   134    75.0     1.2
/home/morama/gas/pili/align/gi-28810263.pep    Begin: 316  End: 638
! gi|28810263|dbj|BAC63201.1| protein...   56    82   134    74.7     1.3
/home/morama/gas/pili/align/orf80.pep    Begin: 49  End: 352
! TRANSLATE of: orf80.seq check: 9824...   45    69   113    70.8     2.1
/home/morama/gas/pili/align/gi-19224137.pep    Begin: 25  End: 342
! gi|19224137|gb|AAL86408.1|AF447492_...   45    69   109    69.8     2.4
/home/morama/gas/pili/align/gi-19224141.pep    Begin: 277  End: 645
! gi|19224141|gb|AAL86412.1|AF447492_...   73    73   118    68.9     2.7
/home/morama/gas/pili/align/gi-21909636.pep    Begin: 44  End: 344
! gi|21909636|ref|NP_663904.1| conser...   45    98    96    66.1     3.8
/home/morama/gas/pili/align/gi-28810259.pep    Begin: 49  End: 349
! gi|28810259|dbj|BAC63197.1| hypothe...   45    98    96    66.0     3.8
/home/morama/gas/pili/align/gas15.pep    Begin: 222  End: 470
! GAS15 GAS15                              42    68    96    63.8       5
/home/morama/gas/pili/align/gi-13621428.pep    Begin: 17  End: 340
! gi|13621428|gb|AAK33238.1| hypothet...   41    41    87    63.6     5.2
/home/morama/gas/pili/align/gi-19224135.pep    Begin: 193  End: 462
! gi|19224135|gb|AAL86406.1|AF447492_...   41    41    86    61.0       7
/home/morama/gas/pili/align/gi-50913505.pep    Begin: 503  End: 966
! gi|50913505|ref|YP_059477.1| Collag...   70   149    86    60.1     7.8
/home/morama/gas/pili/align/gi-13621430.pep    Begin: 60  End: 143
! gi|13621430|gb|AAK33240.1| hypothet...   43    67    67    59.2     8.7
/home/morama/gas/pili/align/gi-19745303.pep    Begin: 44  End: 344
! gi|19745303|ref|NP_606439.1| hypoth...   51   106    69    58.4     9.5
\\End of List


gi-50913506.pep
/home/morama/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes
 MGAS10394]


SCORES   Init1: 3454  Initn: 3454  Opt: 3454  z-score: 1016.2 E(): 4.7e-53
>>/home/morama/gas/pili/align/gi-50913506.pep                (556 aa)
 initn: 3454 init1: 3454 opt: 3454 Z-score: 1016.2 expect(): 4.7e-53
Smith-Waterman score: 3454;   100.0% identity in 556 aa overlap
```

FIGURE 58A

PCT/US05/27239

(1-556:1-556)

```
                10        20        30        40        50        60
gi-50913506. MTNRRETVREKILITAKKLMLACLAILAVVGLGMTRVSALSKDDTAQLKITNIEGGPTVT
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913506. MTNRRETVREKILITAKKLMLACLAILAVVGLGMTRVSALSKDDTAQLKITNIEGGPTVT
                10        20        30        40        50        60

                70        80        90       100       110       120
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSIS
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSIS
                70        80        90       100       110       120

               130       140       150       160       170       180
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYNGEGNLVTKNIDSKSNYLYGQTSV
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYNGEGNLVTKNIDSKSNYLYGQTSV
               130       140       150       160       170       180

               190       200       210       220       230       240
gi-50913506. AKSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFELPSYTKEAVNKTVYVSDNMSEGL
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913506. AKSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFELPSYTKEAVNKTVYVSDNMSEGL
               190       200       210       220       230       240

               250       260       270       280       290       300
gi-50913506. TFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYK
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913506. TFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYK
               250       260       270       280       290       300

               310       320       330       340       350       360
gi-50913506. AVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913506. AVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIA
               310       320       330       340       350       360

               370       380       390       400       410       420
gi-50913506. FRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYS
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913506. FRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYS
               370       380       390       400       410       420

               430       440       450       460       470       480
gi-50913506. LNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDSRPTGN
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913506. LNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDSRPTGN
               430       440       450       460       470       480

               490       500       510       520       530       540
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPSTGSIGTYLFKAIGSA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPSTGSIGTYLFKAIGSA
               490       500       510       520       530       540

               550
gi-50913506. AMIGAIGIYIVKRRKA
             ||||||||||||||||
gi-50913506. AMIGAIGIYIVKRRKA
               550
```

FIGURE 58B

```
gi-50913506.pep
/home/morama/gas/pili/align/orf84.pep

TRANSLATE of: orf84.seq check: 7868 from: 1 to: 2088
generated symbols 1 to: 696.
 GETSEQ from morama, September 13, 2004 17:07.


SCORES   Init1: 57    Initn: 83    Opt: 135   z-score: 75.1  E(): 1.2
>>/home/morama/gas/pili/align/orf84.pep                    (696 aa)
 initn:  83 init1:  57 opt: 135 Z-score: 75.1 expect():  1.2
Smith-Waterman score: 146;    24.4% identity in 262 aa overlap
 (232-462:316-567)

                 210       220       230       240       250
gi-50913506. KTTSLGSVLSYSLTFELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVEWKGKMAN----
                                                |:|:  :|::: ::   |: :|:: :
orf84.pep    EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
               290       300       310       320       330       340

             260       270       280       290       300       310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNISYKAVVNNKAIVGEE
             | :: :: | || :   | ::: :::   :     |  |: :|:|  ::  : | | ||:
orf84.pep    IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNGHIDITEDTTPDI-VSGENQMKQIEGED
                 350       360       370       380       390

                      320       330       340       350
gi-50913506. GNP-------NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ
             ::|      |  ||  :: |     |::| |::: |::| |  :|::|:: ::  :| :
orf84.pep    SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIE
           400       410       420       430       440       450

                360       370       380       390       400       410
gi-50913506. ------IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
                  | | | | ::   | || : : |:|:| |:  | ::| : :  :  ||| : |
orf84.pep    EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
           460       470       480       490       500       510

                420       430       440       450       460       470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
               || || : |  |: :: :  ||:: |:: ::   ::: |  | | :| ||
orf84.pep    TAAPDGYEIATAITFTVNEQGQVTVNGKATKGDAHIVMV-DAYKPTKGSGQVIDIEEKLP
              520       530       540       550       560       570

                480       490       500       510       520       530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPSTGSIG

orf84.pep    DEQGHSGSTTEIEDSKSSDVIIGGQGQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHF
               580       590       600       610       620       630


gi-50913506.pep
/home/morama/gas/pili/align/gi-19745307.pep

gi|19745307|ref|NP_606443.1| protein F2-like protein [Streptococcus pyogenes MGA
S8232]


SCORES   Init1: 57    Initn: 83    Opt: 135   z-score: 75.1  E(): 1.2
>>/home/morama/gas/pili/align/gi-19745307.pep              (696 aa)
 initn:  83 init1:  57 opt: 135 Z-score: 75.1 expect():  1.2
Smith-Waterman score: 146;    24.4% identity in 262 aa overlap
 (232-462:316-567)
```

FIGURE 58C

```
                 210        220        230        240        250
gi-50913506.  KTTSLGSVLSYSLTFELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVEWKGKMAN----
                                                 |:|:  :|::: ::    |: :|:: :
gi-19745307.  EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
                290        300        310        320        330        340

              260        270        280        290        300        310
gi-50913506.  ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNISYKAVVNNKAIVGEE
              | :: :: | || :   | ::: :::   :      |  |: :|:|  ::  : | | ||:
gi-19745307.  IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNGHIDITEDTTPDI-VSGENQMKQIEGED
                  350        360        370        380        390

                       320        330        340        350
gi-50913506.  GNP-------NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ
              ::|      |  ||  :: |     |::| |::: |::| |  :|::|:: ::  :| :
gi-19745307.  SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIE
            400        410        420        430        440        450

                 360        370        380        390        400        410
gi-50913506.  ------IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
                  | | | | ::   | || : : |:|:| |:  |  ::| : :  :  |||: : |
gi-19745307.  EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
            460        470        480        490        500        510

                 420        430        440        450        460        470
gi-50913506.  LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
               || || : |  |: :: :  ||::|:: ::   ::: |  | | :| ||
gi-19745307.  TAAPDGYEIATAITFTVNEQGQVTVNGKATKGDAHIVMV-DAYKPTKGSGQVIDIEEKLP
                520        530        540        550        560        570

                  480        490        500        510        520        530
gi-50913506.  YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPSTGSIG

gi-19745307.  DEQGHSGSTTEIEDSKSSDVIIGGQGQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHF
                580        590        600        610        620        630


gi-50913506.pep
/home/morama/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP_663908.1| protein F2 like fibronectin-binding protein [Strept
ococcus pyogenes MGAS315]


SCORES   init1: 56   initn: 81   opt: 134   z-score: 75.0   E(): 1.2
>>/home/morama/gas/pili/align/gi-21909640.pep              (656 aa)
 initn: 81 init1: 56 opt: 134 z-score: 75.0 expect(): 1.2
Smith-Waterman score: 156;    23.9% identity in 347 aa overlap
 (232-547:202-524)

                 210        220        230        240        250
gi-50913506.  KTTSLGSVLSYSLTFELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVEWKGKMAN----
                                                 |:|:  :|::: ::    |: :|:: :
gi-21909640.  EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
                180        190        200        210        220

              260        270        280        290        300        310
gi-50913506.  ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNISYKAVVNNKAIVGEE
              | :: :: | || :   | ::: :::   :      |  |: :|:|  ::  : | | ||:
gi-21909640.  IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNGHIDITEDTTPDI-VSGENQMKQIEGED
            230        240        250        260        270        280
```

FIGURE 58D

PCT/US05/27239

```
                    320               330         340         350
gi-50913506. GNP--------NKAEFFYSNNP------TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ
             ::|        |  ||  :: |      |::| |::: |::| |  :|::|:: ::  :| :
gi-21909640. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIE
               290        300        310        320        330        340

                360        370        380        390        400        410
gi-50913506. ------IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
                  | | | | ::    | || : : |:|:| |:   |  ::| : :  :  ||| : |
gi-21909640. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
               350        360        370        380        390        400

                420        430        440        450        460        470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
               || || : |  |: :: :  ||:: |:: ::   ::: |  | | :| ||
gi-21909640. TAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIVMV-DAYKPTKGSGQV---------
                 410        420        430        440        450

                480        490        500        510        520        530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPSTGSIG
               || :    | ::::  ||  : |: : :    |:| |:  |:  :|: |    :||
gi-21909640. --IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTE--DTQSGMTGHSGST-
                 460        470        480        490        500

                540        550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
             | :  : :| ::||: |
gi-21909640. TEIEDSKSSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETT
             510        520        530        540        550        560


gi-50913506.pep
/home/morama/gas/pili/align/gi-28810263.pep

gi|28810263|dbj|BAC63201.1| protein F2-like protein [Streptococcus pyogenes SSI-
1]


SCORES   Init1: 56    Initn: 82    Opt: 134   z-score: 74.7  E(): 1.3
>>/home/morama/gas/pili/align/gi-28810263.pep                (733 aa)
 initn:  82 init1:  56 opt: 134 Z-score: 74.7 expect():  1.3
Smith-Waterman score: 155;    23.6% identity in 347 aa overlap
 (232-547:316-638)

                210        220        230        240        250
gi-50913506. KTTSLGSVLSYSLTTFELPSYTKEAVNKTVYVSDNMSEGLTTNENSLTVEWKGKMAN----
                                                |:|:  :|:: ::  |: :|:: :
gi-28810263. EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
               290        300        310        320        330        340

             260        270        280        290        300        310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNISYKAVVNNKAIVGEE
             | :: :: | ||  :   | :::  :::   :     |  |: :|:|  ::  : | | ||:
gi-28810263. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNGHIDITEDTTPDI-VSGENQMKQIEGED-
               350        360        370        380        390

                320        330        340        350
gi-50913506. GNP-------NKAEFFYSNNP------TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ
             ::|        |  ||  :: |      |::| |::: |::| |  :|::|:: ::  :| :
gi-28810263. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIE
           400        410        420        430        440        450

                360        370        380        390        400        410
```

FIGURE 58E

PCT/US2005/027239

```
gi-50913506. ------IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
                   | | | | ::   | || : : |:|:| |:  |  ::| : :  :  |||  : |
gi-28810263. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
           460       470       480       490       500       510

                  420       430       440       450       460       470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
               || || : |  |: :: :  ||:: |:: ::   ::: |  | | :| ||
gi-28810263. TAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIVMV-DAYKPTKGSGQV--------
              520       530       540       550       560

                  480       490       500       510       520       530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPSTGSIG
               || :    | ::::  ||  : |: : :   |:| |:  |:  :|: |    :||
gi-28810263. --IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTE--DTQSGMTGHSGST-
               570       580       590       600       610       620

                  540       550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
               | :  : :| :::|: |
gi-28810263. TKIEDSKSSDVIVGGQGQIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNKEPESNS
                  630       640       650       660       670       680


gi-50913506.pep
/home/morama/gas/pili/align/orf80.pep

TRANSLATE of: orf80.seq check: 9824 from: 1 to: 1056
generated symbols 1 to: 352.
 GETSEQ from morama, September 13, 2004 17:11.


SCORES   Init1: 45    Initn: 69    Opt: 113   z-score: 70.8  E(): 2.1
>>/home/morama/gas/pili/align/orf80.pep                    (352 aa)
 initn:  69 init1:  45 opt: 113 Z-score: 70.8 expect():  2.1
Smith-Waterman score: 123;    22.8% identity in 311 aa overlap
 (284-556:49-352)

               260       270       280       290       300       310
gi-50913506. KMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVN-NKAIVGEE
                                          :: ||: | : |: ::  ::: : :  |:|
orf80.pep    ATALGTASLNQNVKAETAGVVTGKSLQVTKTMTYDDEEVLMPETAFTFTIEPDMTASGKE
             20        30        40        50        60        70

                  320       330       340       350       360       370
gi-50913506. GNPN-KAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIAFRKVDSVSKTP
             |: : |  : : :       | | ||  :| : |:: : ||: :   :| : | .:
orf80.pep    GSLDIKNGIVEGLDKQVTVKYKNTDKTSQKTK-IAQFDFSKVKFPAIGVYRYMVSEKNDK
             80        90        100       110       120       130

                  380       390       400       410       420
gi-50913506. LIGAIFGVYDTSNKLIDIVTTNKNG--------YAISTQ-VSSGKYKIKELKAPKGYSLN
               |   :|| :.:  :|: : || :       | :| : :|| |  |:  :: |  ||:
orf80.pep    KDGI---TYDDKKWTVDVYVGNKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLK
             140       150       160       170       180       190

                  430               440       450       460
gi-50913506. TETYEITANW---------VTATVKTSANSKSTTYTSDKNKATDNSEQVG-----WLKNGI
               | :||:|           | |:: |   |: : :: :: :: ::  :|       | :|
orf80.pep    IEK-QITGNAGDRKKSFNFTLTLQPSEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGK
                 200       210       220       230       240       250

           470       480                 490       500       510
```

FIGURE 58F

PCT/US05/27239

```
gi-50913506.  FYSIDSRPTGNDV--------KEAYI------ESTKALTDGTTFSKSNEGSGTVLLETDI
                :::  | | :          |::|        |: |  ::  |:| :|: :      |  :
orf80.pep     SVMLSKLPIGINYYLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQNQKTDESADEIVV
                  260       270        280        290        300        310

                520        530        540        550
gi-50913506.  PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
               | :  ::| || :||    |: | : ||:: |||:||:||
orf80.pep     TNKRDTQVP-TGVVGTLAPFAVLSIVAIGGV-IYITKRKKA
                  320        330        340        350


gi-50913506.pep
/home/morama/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]


SCORES   Init1: 45    Initn: 69    Opt: 109   z-score: 69.8  E(): 2.4
>>/home/morama/gas/pili/align/gi-19224137.pep              (342 aa)
 initn:  69 init1:  45 opt: 109 Z-score: 69.8 expect():  2.4
Smith-Waterman score: 169;    26.0% identity in 334 aa overlap
 (257-556:25-342)

             230        240        250        260        270        280
gi-50913506.  NKTVYVSDNMSEGLTFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFI
                                                |:  : : :| : :: | | ::| ||
gi-19224137.          MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSITN-FN----
                             10         20         30         40

             290        300        310        320        330        340
gi-50913506.  YDSLESISPNISYKAVVN-NKAIVGEEGN-PNKAEFFYSNNPTKGNTYDNLDKKPDKGNG
               |:|  : :|: :|    || ::| :| |:| | |  :  :|:  |  :|:| ||   | :
gi-19224137.  DDTL--LMPKTDYTFSVNPDSAATGTESNLPIKPGIAVNNQDIK-VSYSNTDKTSGKEKQ
            50         60         70         80         90         100

               350        360        370        380        390        400
gi-50913506.  ITSKEDSKIVYTYQIAFRKVDSVSKTPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQV
              ::  :  |:::     :| | : :|    |:  || || :: |:|: : |::  ::  :
gi-19224137.  VVV-DFMKVTFPSVGIYRYVVTENK----GTAEGVTYDDTKWLVDVYVGNNEKGGLEPKY
                110        120        130            140        150        160

              410        420        430        440        450
gi-50913506.  SSGKYKIKELKAPKGY--SLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNS--
               :|  :  | | :  |::| : :|  : ||:::    :: : | | : |:  : |
gi-19224137.  IVSKKGDSATKEPIQFNNSFETTSLKIEKE-VTGNTGDHKKAFTFTLTLQPNEYYEASSV
                170        180        190        200        210        220

              460        470                480                490
gi-50913506.  ---EQVGWLKN---GIFYSI---DSR--------PTGND--VKEAYIE-----STKALTDG
                |: | |:  | |:: ||:      |:| :  |:|| :     :| :| ||
gi-19224137.  VKIEENGQTKDVKIGEAYKFTLNDSQSVILSKLPVGINYKVEEAEANQGGYTTTATLKDG
                   230        240        250        260        270        280

              500        510        520        530        540        550
gi-50913506.  TTFSKSNEG----SGTVLLETDIPNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVK
               :|  | |    :  :  |  : |::  ::| || :||    |: | : ||:: |||:|
gi-19224137.  EKLSTYNLGQEHKTDKTADEIVVTNNRDTQVP-TGVVGTLAPFAVLSIVAIGGV-IYITK
                   290        300        310        320        330

gi-50913506.  RRKA
              |:||
```

FIGURE 58G

```
gi-19224137. RKKA
             340


gi-50913506.pep
/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]


SCORES   Init1: 73    Initn: 73    Opt: 118   z-score: 68.9  E(): 2.7
>>/home/morama/gas/pili/align/gi-19224141.pep            (1161 aa)
 initn:  73 init1:  73 opt: 118 Z-score: 68.9 expect():  2.7
Smith-Waterman score: 174;    23.9% identity in 406 aa overlap
 (115-483:277-645)

                 90        100        110        120        130        140
gi-50913506. VSLTETGPTSQEITTIANGINTGKIKPFSTENVSISNGTATYNARGASVY--IALLTGAT
                                                 ||:|||| :| ::: :    : :| |
gi-19224141. IYTFTDYIAGLDKVQLSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTVNVLYGNE
              250        260        270        280        290        300

                150        160        170        180        190
gi-50913506. DGRTYNPILLAASYNGEGNLVTKNIDSKSNYLYGQTSVAKSSLPSITKKVTG-------T
             : :  | |  : |  | |:: : | ::     |  ::  ::::|  | :: |       |
gi-19224141. STKESNYITNGLSNVG-GSIESYNTETGEFVWYVYVNPNRTNIPYATMNLWGFGRARSNT
              310        320        330        340        350        360

            200        210        220        230        240        250
gi-50913506. ID---DVNKKTTSLGSVLSYSLTF--ELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVE
             |   |:| ::: || :  | :     :|||      |:| :  :| :: ||  :|:
gi-19224141. SDLENDANTSSAELGEIQVYEVPEGEKLPSSYGVDVTKLTLRTD-ITAGLGNGFQ-----
              370        380        390        400        410

               260        270        280        290        300        310
gi-50913506. WKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVG
                        |::  :| ::    ||  | :||     : : : | | :| :: :::
gi-19224141. ---------------MTKRQRIDFG---NNIQNKAFIIKV--TGKTDQSGKPLVVQSNLAS
                          420        430        440        450        460

                320        330        340        350                360
gi-50913506. EEGNPNKAEFFYSNNPTKGNTY--DNLDKKPDKGNGITSKEDSKIVYTY-------QIAF
             :|  : | |     :|: ||:|  ::::  :|:||:|  ::| :|   |          |: |
gi-19224141. FRGASEYAAF-----TPVGGNVYFQNEIALSPSKGSGSGKSEETKPSITVANLKRVAQLRF
                   470        480        490        500        510

               370        380        390        400        410        420
gi-50913506. RKVDSVSKTPLIGAIFGVYDTSNKLIDI-VTTNKNGYAISTQVSSGKYKIKELKAPKGYS
             :|: |:::::||  | | : :::::  : :::| :| :  :::|| | : | ||||||:
gi-19224141. KKM-STDNVPLPEAAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQ.
              520        530        540        550        560        570

                          430        440        450        460
gi-50913506. -----LNTETYEIT------ANWVT--ATVKTSANSKSTTYTSDKNKATDNSEQVGWLKN
                  | | | : |      ::| :  :::||: || | :| :: |:  | :::::: | ::
gi-19224141. QVTEKLATVTVDTTKPAEEMVTWGSPHSSVKVEAN-KEVTIVNHKETLTFSGKKI-WEND
              580        590        600        610        620        630

            470        480        490        500        510        520
gi-50913506. GIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPSTG
                  |:||:  :|:
gi-19224141. ----RPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQEYKYSVEE
                    640        650        660        670        680
```

FIGURE 58H

```
gi-50913506.pep
/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyoge
nes MGAS315]


SCORES   Init1: 45     Initn: 98     Opt: 96     z-score: 66.1  E(): 3.8
>>/home/morama/gas/pili/align/gi-21909636.pep              (344 aa)
 initn:  98 init1:  45 opt:  96 Z-score: 66.1 expect():  3.8
Smith-Waterman score: 181;    25.3% identity in 312 aa overlap
 (298-556:44-344)

              270       280       290       300       310       320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
                                              :: : ::|:::: :      |:|   ::: :
gi-21909636. ATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVE---ADSTA
                   20        30        40        50        60          70

              330       340                 350       360
gi-50913506. KGNTYDNLDKKPDKGNGIT----------SKEDSKIVYTYQIAFRKV----------DSVSK
             :|:| |:|: ||   ||:|          :| |||:  | :: | ||          :||:
gi-21909636. SGKTKDGLEIKPGIVNGLTEQIISYTNTDKPDSKVKST-EFDFSKVVFPGIGVYRYTVSE
                        80        90       100       110       120

           370       380       390       400       410       420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY
               | : |: |||:: :|: : ||:| :: :  :| :  ::| | .:  |: | :
gi-21909636. KQ--GDVEGITYDTKKWTVDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSL
           130       140       150       160       170       180

             430       440               450       460
gi-50913506. EITANWVTATVKTSAN-------SKSTTYTSDK----NKATDNSE-QVGW-----LKNGI
             ::  |    | : : :        ::||:: :|:    :|:::: | ::|       ||||
gi-21909636. KVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQKGNEKFEVKIGTPYKFKLKNGE
             190       200       210       220       230       240

           470       480       490       500       510
gi-50913506. FYSIDSRPTGNDVKEAYIESTKALTDGTTFSKS-NEGSGTVLL-------ETD-------
               ::|: |:|   |   :|::|   ||   : | :||:|   :        :||
gi-21909636. SIQLDKLPVGITYKVNEMEANK---DGYKTTASLKEGDGQSKMYQLDMEQKTDESADEIV
             250       260       270       280       290       300

               520       530       540       550
gi-50913506. IPNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
             : | :  ::| || :||  . |: | : ||:: |||:||:||
gi-21909636. VTNKRDTQVP-TGVVGTLAPFAVLSIVAIGGV-IYITKRKKA
               310       320       330       340


gi-50913506.pep
/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]


SCORES   Init1: 45     Initn: 98     Opt: 96     z-score: 66.0  E(): 3.8
>>/home/morama/gas/pili/align/gi-28810259.pep              (349 aa)
 initn:  98 init1:  45 opt:  96 Z-score: 66.0 expect():  3.8
Smith-Waterman score: 181;    25.3% identity in 312 aa overlap
 (298-556:49-349)
```

FIGURE 58I

PCT/US05/27239

```
             270       280       290       300       310       320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
                                                :: : ::|:::: :     |:|   ::: :
gi-28810259. ATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVE---ADSTA
             20        30        40        50        60        70

             330       340                 350       360
gi-50913506. KGNTYDNLDKKPDKGNGIT---------SKEDSKIVYTYQIAFRKV---------DSVSK
             :|:| |:|: ||    ||:|          :| |||:  | :: | ||          :||:
gi-28810259. SGKTKDGLEIKPGIVNGLTEQIISYTNTDKPDSKVKST-EFDFSKVVFPGIGVYRYTVSE
                80        90        100       110       120       130

          370       380       390       400       410       420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY
               | : |: |||:: :|: : ||:| :: :   :| :  ::| | ::  |: | :
gi-28810259. KQ--GDVEGITYDTKKWTVDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSL
                  140       150       160       170       180       190

             430       440               450               460
gi-50913506. EITANWVTATVKTSAN--------SKSTTYTSDK----NKATDNSE-QVGW-----LKNGI
             ::  |     | : : :         ::||:: :|:     :|:::: | ::|      ||||
gi-28810259. KVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQKGNEKFEVKIGTPYKFKLKNGE
                  200       210       220       230       240       250

          470       480       490       500       510
gi-50913506. FYSIDSRPTGNDVKEAYIESTKALTDGTTFSKS-NEGSGTVLL-------ETD-------
               ::|: |:|   |   :|::|   ||   : | :||:|   :       :||
gi-28810259. SIQLDKLPVGITYKVNEMEANK---DGYKTTASLKEGDGQSKMYQLDMEQKTDESADEIV
                  260       270       280       290       300

             520       530       540       550
gi-50913506. IPNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
             : | :  ::| || :||     |: | : ||:: |||:||:||
gi-28810259. VTNKRDTQVP-TGVVGTLAPFAVLSIVAIGGV-IYITKRKKA
          310       320       330       340
```

gi-50913506.pep
/home/morama/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Init1: 42 Initn: 68 Opt: 96 z-score: 63.8 E(): 5
>>/home/morama/gas/pili/align/gas15.pep (762 aa)
initn: 68 init1: 42 opt: 96 z-score: 63.8 expect(): 5
Smith-Waterman score: 96; 23.4% identity in 269 aa overlap
(283-535:222-470)

```
                260       270       280       290       300
gi-50913506. GKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLES-ISPNISYKA---VVNN--K
                                              ||:: ::|:: |:||:: |    | ::
gas15.pep    VWYYSDNAPISNPDESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPKQVPDDFQL
                  200       210       220       230       240       250

             310       320       330       340       350       360
gi-50913506. AIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIAFRK--V
             :|    | : :| :  |:|  : |:      | |  |:      :: :   | :: :||  :
gas15.pep    SIFESEDKGDKYNKGYQNLLSGGLVPT---KPPTPGDPPMPPNQPQ---TTSVLIRKYAI
                  260       270           280       290           300

               370       380       390       400       410       420
```

FIGURE 58J

PCT/US05/27239

```
gi-50913506.  DSVSKTPLIGAIFGVY-DTSNKL-IDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYSLN
gas15.pep     GDYSKL-LEGATLQLTGDNVNSFQARVFSSNDIGERI--ELSDGTYTLTELNSPAGYSIA
                310       320       330       340       350       360

                  430       440       450       460       470
gi-50913506.  TE-TYEITANWVTATV--KTSANSKSTTYTSDKNKATDNSEQVGWLKN---GIFYSIDSR
gas15.pep     EPITFKVEAGKVYTIIDGKQIENPNKEIVEPYSVEAYNDFEEFSVLTTQNYAKFYYAKNK
                370       380       390       400       410       420

                480       490       500       510       520       530
gi-50913506.  PTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPSTGSIGTYLFKA
gas15.pep     NGSSQVVYCFNADLKSPPD------SEDGGKTMT-----PDFTTGEVKYTHIAGRDLFKY
                430       440             450             460       470

                540       550
gi-50913506.  IGSAAMIGAIGIYIVKRRKA

gas15.pep     TVKPRDTDPDTFLKHIKKVIEKGYREKGQAIEYSGLTETQLRAATQLAIYYFTDSAELDK
                      480       490       500       510       520       530


gi-50913506.pep
/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]


SCORES  Init1: 41    Initn: 41    Opt: 87    z-score: 63.6  E(): 5.2
>>/home/morama/gas/pili/align/gi-13621428.pep              (340 aa)
 initn:  41 init1:  41 opt:  87 Z-score: 63.6 expect():  5.2
Smith-Waterman score: 109;    22.6% identity in 345 aa overlap
 (256-556:17-340)

                230       240       250       260       270       280
gi-50913506.  VNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSF
gi-13621428.                                  MKLRHLLLTGAALTSFAATTVHGETVVNGAKLTVTK------NLDL
                                        10        20        30          40

                290       300       310       320       330       340
gi-50913506.  IYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGI
gi-13621428.  VNSN--ALIPNTDFTFKIEPDTTVNEDGNKFKGVAL--NTPMTKVTYTNSDK--GGSNTK
                      50        60        70        80        90

                350               360                       370
gi-50913506.  TSKED-SKI------VYTYQIAFRKVDSV-------------------SKTPLIGAIF
gi-13621428.  TAEFDFSEVTFEKPGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVATYIV
                100       110       120       130       140       150

                380       390       400       410       420       430
gi-50913506.  GVYDTSNKLIDIVTT-NKNGYAISTQVS-SGKYKIKELKAPKGYSLNTETYEITANWVTA
gi-13621428.  GYKEGSKVPIQFKNSLDSTTLTVKKKVSGTGGDRSKDFNF--GLTLKANQYYKASEKVMI
                160       170       180       190       200       210

                440       450       460       470       480       490
gi-50913506.  TVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDSRPTGND--VKEAYIESTKAL
```

FIGURE 58K

```
gi-13621428. E-KTTKGGQAPVQT----EASIDQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYKSEKYT
                  220          230        240        250        260

                           500        510        520        530        540
gi-50913506. T--------DGTT-----FSKSNEGSGTVLLETDIPNTKLGELPSTGSIGTYLFKAIGSAA
             |       ||::     |  :| |    :   : |. |  |:|:  :: :  : |:| :|
gi-13621428. TNVEVSPQDGAVKNIAGNSTEQETSTDKDMTITFTNKKDFEVPTGVAMTVAPYIALGIVA
             270        280        290        300        310        320

                 550
gi-50913506. MIGAIGIYIVKRRKA
             : ||:  |:||:::|
gi-13621428. VGGAL--YFVKKKNA
             330        340


gi-50913506.pep
/home/morama/gas/pili/align/gi-19224135.pep

gi|19224135|gb|AAL86406.1|AF447492_3 Cpa [Streptococcus pyogenes]


SCORES  Init1: 41    Initn: 41    Opt: 86    z-score: 61.0  E(): 7
>>/home/morama/gas/pili/align/gi-19224135.pep              (756 aa)
 initn:  41 init1:  41 opt:  86 Z-score: 61.0 expect():    7
Smith-Waterman score: 101;    19.6% identity in 306 aa overlap
 (243-535:193-462)

                  220        230        240        250        260        270
gi-50913506. SLTFELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANITEDGSVMVENTKIG
                                                :::::: | :::    :||::  |::|:
gi-19224135. PKNANGYMDKIEPLNAILVTQQAVWYYSDSSYGNIKTLWASEL-----KDGKIDFEQVKL-
                  170        180        190        200            210

                  280        290        300        310        320        330
gi-50913506. IAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPN--KAEFFYSNNPTKGN
              :|: :    ::| |:|| | |   ::   |  :  ::: |  :||:    : |: |:
gi-19224135. -MREAYS----KLISDDLEETSKNKLPQGSKLNIFVPQDKSVQNLLSAEYVPESPPAPGQ
             220            230        240        250        260        270

                    340        350        360        370        380
gi-50913506. TYDNLDKKPDKGNGITSKEDSKIVYTY-QIAFRKVDSVSKTPLIGAIFGVYDTSNKLIDI
             :    : |     : :|: | |:  | :  : |:   :   | |    : | ::|   :
gi-19224135. S----PEPP-----VQTKKTSVIIRKYAEGDYSKLLEGATLRLTGE--DILDFQEK---V
                       280        290        300        310

          390        400        410        420        430        440
gi-50913506. VTTNKNGYAISTQVSSGKYKIKELKAPKGYSLNTET-YEITANWVTATVKTSANSKSTT-
             :| :|  | ::|:| | : | ::| ||::  ( :::: : | : | ::: :: :
gi-19224135. FQSNGTGEKI--ELSNGTYTLTETSSPDGYKIAEPIKFRVVNKKVFIVQKDGSQVENPNK
            320        330        340        350        360        370

                  450        460        470        480        490
gi-50913506. -----YTSDKNKATDNSEQVG---WLKNGIFYSIDSRPTGNDVKEAYIESTKALTDGTTF
                |: : :  ::|: ::   :  | ||   ::  :::|  :       :|  :
gi-19224135. EVAEPYSVEAYSDMQDSNYINPETFTPYGKFYYAKNKDKSSQVVYCFN------ADLHSP
             380        390        400        410        420            430

           500        510        520        530        540        550
gi-50913506. SKSNEGSGTVLLETDIPNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
             :|::|:||:  : || : |  |:  | : |: |||
gi-19224135. PESEDGGGTI--DPDISTMK--EVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKG
                  440        450        460        470        480
```

FIGURE 58L

```
gi-19224135. YNKKGDSYNGLTETQFRAATQLAIYYFTDSTDLKTLKTYNNGKGYHGFESMDEKTLAVTK
              490       500       510       520       530       540


gi-50913506.pep
/home/morama/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M
GAS10394]


SCORES   Init1: 70    Initn: 149   Opt: 86    z-score: 60.1  E(): 7.8
>>/home/morama/gas/pili/align/gi-50913505.pep            (1036 aa)
 initn: 149 init1:  70 opt:  86 Z-score: 60.1 expect():  7.8
Smith-Waterman score: 120;    21.5% identity in 469 aa overlap
 (33-428:503-966)

                 10        20        30        40        50        60
gi-50913506. NRRETVREKILITAKKLMLACLAILAVVGLGMTRVS-ALSKDDTAQLKITNIEGGPTVTL
                                                 |:: || :: :||::: :: :::|      |
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNG-----L
             480       490       500       510       520

                 70        80        90       100       110
gi-50913506. YKIGEGVYNTNGDSFINFK----YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSTEN
              :   ::   |  | |:|:    |  |   : :| :     :: |: ::  :| :||::
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS
             530       540       550       560       570       580

                120       130       140               150       160
gi-50913506. VSISNGTATYNARGASVYIALLTGAT--------DGRTYNPILLAAS-----YNGEGNLVTK
             :  |  :: :  |:: : ||| :          |||    |::: :    | || :  :
gi-50913505. LLDPNTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSG
             590       600       610       620       630       640

                170       180               190       200       210
gi-50913506. NIDS--KSNYLYGQ--TSVA----KSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFE
             | :|  ::|   :|  :::|    |:  |:::    |:  |:|: |:| : |:  |: |
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGE
             650       660       670       680       690       700

               220               230               240       250       260
gi-50913506. LPSY-------TKEAVNKTVY--------VSDNMSEGLTF--NFNSLTVEWKGKMANITE
             |          ::::|| |          :||::|: : :  :  :: |  |:|: : ||
gi-50913505. EHYYGITDTAELEKTLNKLVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETE
             710       720       730       740       750       760

                                 270       280       290
gi-50913506. ----------------DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
                             |  |:: :|      :|:  |: :: |:    ::| |:
gi-50913505. ILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDE
             770       780       790       800       810       820

               300       310       320       330       340       350
gi-50913506. SYKAVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDK-GNGITSKEDSK----
             :|:   :|:: :| |: :      :::  ||:  :| | : :  ::|  :|   |
gi-50913505. AYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVI
             830       840       850       860       870       880

                360       370       380       390       400
gi-50913506. IVYTYQIAFRKVDS-VSKTPLIGAIFGVYDTSNKLI-DIVTTNKNGYAISTQVSSGK-YK
              | |  |:| |||:  ::  | |: | :   ::|:: :  ||::||      :::|| |
```

FIGURE 58M

PCT/US05/27239

```
gi-50913505. QVKTVPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYY
             890       900       910       920       930       940

          410       420       430       440       450       460
gi-50913506. IKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGI
             : | ||  ||:|  : :|::
gi-50913505. LYETKAKLGYTLPENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGR
             950       960       970       980       990       1000


gi-50913506.pep
/home/morama/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]


SCORES  Init1: 43    Initn: 67    Opt: 67    z-score: 59.2  E(): 8.7
>>/home/morama/gas/pili/align/gi-13621430.pep              (215 aa)
 initn:  67 init1:  43 opt:  67 Z-score: 59.2 expect():  8.7
Smith-Waterman score: 67;    27.8% identity in 90 aa overlap
 (433-522:60-143)

                410       420       430       440       450       460
gi-50913506. VSSGKYKIKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQV
                                              : |  | |   :|:| || |:::  : ||:
gi-13621430. TASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVT-TSVKGNGKTSFEQL
             30        40        50        60        70        80

                470       480       490       500       510       520
gi-50913506. GWLKNGIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGE
              : : | ::    :  |:: :  | |::  ::  : :   || ||:  |||:: ::||||
gi-13621430. TFSEVGQYHYKIHQLLGKNSQYHYDETVYEVVIYVLY---NEQSGA--LETNLVSNKLGE
             90        100       110       120       130       140

                530       540       550
gi-50913506. LPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA

gi-13621430. TEKSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATIT
                150       160       170       180       190       200


gi-50913506.pep
/home/morama/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP_606439.1| hypothetical protein [Streptococcus pyogenes MGAS82
32]


SCORES  Init1: 51    Initn: 106   Opt: 69    z-score: 58.4  E(): 9.5
>>/home/morama/gas/pili/align/gi-19745303.pep              (344 aa)
 initn: 106 init1:  51 opt:  69 Z-score: 58.4 expect():  9.5
Smith-Waterman score: 129;    24.0% identity in 308 aa overlap
 (298-556:44-344)

             270       280       290       300       310       320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
                                              :: : :::|::: :      |:|   ::: :
gi-19745303. ATALGTASLNQNVKAETAGVIDGSTLVVKKTFPSYTDDKVLMPKADYTFKVE---ADDNA
                  20        30        40        50        60        70

             330       340                     350       360       370
gi-50913506. KGNTYDNLDKKPDKGNGI--------------TSKEDSKIVYTYQIAFRKVDSVSKT--P
             ||:|  |:|| ||   :|:              |:|| |      :: |  |      |
```

FIGURE 58N

```
gi-19745303. KGKTKDGLDIKPGVIDGIENTKTIHYGNSDKTTAKEKSVNFDFANVKFPGVGVYRYTVSE
                    80        90        100       110       120       130

                380       390       400       410       420
gi-50913506. LIGAIFGV-YDTSNKLIDIVTTNKN--GYAISTQVSS-GKYKIKELKAPKGYSLNTETYE
             : |   |: ||::: :|: ::|::  |: : ||: |  : |:    |:: ::| : :
gi-19745303. VNGNKAGIAYDSQQWTVDVYVVNREDGGFEAKYIVSTEGGQSDKKPVLFKNF-FDTTSLK
                140       150       160       170       180

             430       440       450       460       470
gi-50913506. ITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF--------YSIDSR---
             :| : ||::: |  | |  |: :::: |: |::|          ||:  :
gi-19745303. VTKK-VTGNTGEHQRSFSFTLLLTPNECFEKGQVVNILQGGETKKVVIGEEEYSFTLKDKE
          190       200       210       220       230       240

                          480       490       500       510
gi-50913506. --------PTG-------NDVKEAYIESTKALTDGTTFSKSNEG-SGTVLLETD---IPNT
                   |:|       :|| :  ::: :| || : :  | | | |:  ||  : |
gi-19745303. SVTLSQLPVGIEYKVTEEDVTKDGYKTSATLKDGDVTDGYNLGDSKTTDKSTDEIVVTNK
             250       260       270       280       290       300

             520       530       540       550
gi-50913506. KLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
             :  ::| || :||    |: | : ||:: |||:||:||
gi-19745303. RDTQVP-TGVVGTLAPFAVLSIVAIGGV-IYITKRKKA
             310       320       330       340


! Distributed over 1 thread.
!     Start time: Wed Sep 15 18:45:54 2004
! Completion time: Wed Sep 15 18:46:02 2004

! CPU time used:
!       Database scan:  0:00:00.1
! Post-scan processing:  0:00:01.9
!      Total CPU time:  0:00:02.0
! Output File: gi-50913506.fasta
```

FIGURE 580

PCT/US05/27239

!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-13621430.pep from: 1 to: 215 September 15, 2004 18:45

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp
Variable pamfactor used
Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:
Each histogram symbol represents 1 search set sequences
z-scores computed from opt scores

z-score	obs (=)	exp (*)	
< 20	0	0	:
22	0	0	:
24	0	0	:
26	0	0	:
28	0	0	:
30	0	0	:
32	0	0	:
34	0	1	:*
36	0	2	: *
38	0	3	: *
40	0	4	: *
42	0	5	: *
44	1	5	:= *
46	7	5	:====*==
48	21	5	:====*================
50	7	5	:====*==
52	2	4	:== *
54	4	3	:==*=
56	3	3	:==*
58	0	2	: *
60	4	2	:=*==
62	1	2	:=*
64	0	1	:*
66	0	1	:*
68	0	1	:*
70	0	1	:*
72	0	0	:
74	0	0	:
76	0	0	:
78	0	0	:
80	3	0	:===
82	2	0	:==
84	0	0	:
86	0	0	:
88	0	0	:
90	0	0	:
92	0	0	:
94	0	0	:
96	0	0	:
98	0	0	:
100	0	0	:

FIGURE 59

PCT/US05/27239

```
 102       0       0:
 104       0       0:
 106       0       0:
 108       0       0:
 110       0       0:
 112       0       0:
 114       0       0:
 116       0       0:
 118       0       0:
>120       1       0:=

Joining threshold: 36, opt. threshold: 24, opt. width:  16, reg.-scaled


The best scores are:                                  init1 initn   opt    z-sc E(55)..

/home/morama/gas/pili/align/gi-13621430.pep    Begin: 1  End: 215
! gi|13621430|gb|AAK33240.1| hypothet... 1338  1338  1338   233.9  1.8e-09
/home/morama/gas/pili/align/gi-19745305.pep    Begin: 1  End: 193
! gi|19745305|ref|NP_606441.1| hypoth...  163   243   273    82.2    0.5
/home/morama/gas/pili/align/gi-28810261.pep    Begin: 2  End: 187
! gi|28810261|dbj|BAC63199.1| hypothe...  164   239   268    81.5   0.55
/home/morama/gas/pili/align/gi-19224139.pep    Begin: 2  End: 187
! gi|19224139|gb|AAL86410.1|AF447492_...  164   236   265    81.0   0.57
/home/morama/gas/pili/align/orf82.pep    Begin: 30  End: 222
! TRANSLATE of: orf82.seq check: 4296...  163   235   264    81.0   0.58
/home/morama/gas/pili/align/gi-21909638.pep    Begin: 2  End: 181
! gi|21909638|ref|NP_663906.1| hypoth...  164   239   261    80.5   0.62
/home/morama/gas/pili/align/gi-19745303.pep    Begin: 84  End: 183
! gi|19745303|ref|NP_606439.1| hypoth...  121   121   126    61.4    6.7
/home/morama/gas/pili/align/gi-13621428.pep    Begin: 6  End: 174
! gi|13621428|gb|AAK33238.1| hypothet...   58    86   122    60.9    7.2
/home/morama/gas/pili/align/gi-19224137.pep    Begin: 93  End: 201
! gi|19224137|gb|AAL86408.1|AF447492_...   88    88   119    60.4    7.5
/home/morama/gas/pili/align/gi-50913503.pep    Begin: 549  End: 625
! gi|50913503|ref|YP_059475.1| Fibron...   73    73   117    60.4    7.6
/home/morama/gas/pili/align/gi-19224134.pep    Begin: 631  End: 697
! gi|19224134|gb|AAL86405.1|AF447492_...   73    73   115    60.1    7.8
\\End of List


gi-13621430.pep
/home/morama/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]


SCORES  Init1: 1338  Initn: 1338  Opt: 1338  z-score: 233.9 E(): 1.8e-09
>>/home/morama/gas/pili/align/gi-13621430.pep              (215 aa)
 initn: 1338 init1: 1338 opt: 1338 z-score: 233.9 expect(): 1.8e-09
Smith-Waterman score: 1338;   100.0% identity in 215 aa overlap
 (1-215:1-215)

                    10        20        30        40        50        60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
                    10        20        30        40        50        60

                    70        80        90       100       110       120
gi-13621430. EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-13621430. EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
```

FIGURE 59A

PCT/US05/27239

```
                    70        80        90       100       110       120

                   130       140       150       160       170       180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGI
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGI
                   130       140       150       160       170       180

                   190       200       210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
             ||||||||||||||||||||||||||||||||||
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
                   190       200       210


gi-13621430.pep
/home/morama/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP_606441.1| hypothetical protein [Streptococcus pyogenes MGAS82
32]


SCORES   Init1: 163   Initn: 243   Opt: 273   z-score: 82.2  E(): 0.5
>>/home/morama/gas/pili/align/gi-19745305.pep              (195 aa)
 initn: 243 init1: 163 opt: 273 Z-score: 82.2 expect():  0.5
Smith-Waterman score: 320;    31.9% identity in 213 aa overlap
 (1-213:1-193)

                    10        20        30        40        50        60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
             |:|    :::: ::|  : : ::| |:: |:: :| | | :: | :: |      | ::
gi-19745305. MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTS------FSVAL
                    10        20        30        40              50

                    70        80        90       100       110       120
gi-13621430. EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
             |::|  : : : :|  : |:||:||  |||: |||| |:::|  ::|::|: | ||::|:
gi-19745305. ESIDAMKTI-DEIT--IAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
                60        70        80        90       100       110

                   130       140       150       160       170       180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGI
             :|| |:|: |:| ::::| : |: ||| : || :   |  |:||| |:
gi-19745305. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPDI-----PKTP----
                  120       130       140       150             160

                   190       200       210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
             || :||: | :: |:|||:: ::| : ||||:
gi-19745305. LPLAGEVKSLLGILSIVLLGLLVLLYV-KKLKSRL
                  170       180       190


gi-13621430.pep
/home/morama/gas/pili/align/gi-28810261.pep

gi|28810261|dbj|BAC63199.1| hypothetical protein [Streptococcus pyogenes SSI-1]


SCORES   Init1: 164   Initn: 239   Opt: 268   z-score: 81.5  E(): 0.55
>>/home/morama/gas/pili/align/gi-28810261.pep              (189 aa)
 initn: 239 init1: 164 opt: 268 Z-score: 81.5 expect(): 0.55
Smith-Waterman score: 306;    30.6% identity in 206 aa overlap
```

FIGURE 59B

(8-213:2-187) US05/27239

```
                    10         20         30         40         50         60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
                 :::: ::|  : : ::| |:: |:: :| | | :: | ::          | :::
gi-28810261.       MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP------FSIAL
                          10         20         30         40

                    70         80         90        100        110        120
gi-13621430. EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
             |::|  : : :    :: |:||:||  |||: |||| |:::|  ::|::|: | ||::|:
gi-28810261. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
             50         60         70         80         90        100

                   130        140        150        160        170        180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGI
             :|| |:|: |:| ::::| : |: ||| : || ::  |   |:||:       |:
gi-28810261. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNI-----PKTP----
               110        120        130        140        150

                   190        200        210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
             || :||: | :: |:|||:: ::|  : ||||:
gi-28810261. LPLAGEVKSLLGILSIVLLGLLVLLYV-KKLKSRL
               160        170        180


gi-13621430.pep
/home/morama/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]


SCORES   Init1: 164   Initn: 236   Opt: 265   z-score: 81.0  E(): 0.57
>>/home/morama/gas/pili/align/gi-19224139.pep              (189 aa)
 initn: 236 init1: 164 opt: 265 Z-score: 81.0 expect(): 0.57
Smith-Waterman score: 303;    30.6% identity in 206 aa overlap
 (8-213:2-187)

                    10         20         30         40         50         60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
                 :::: ::|  : : ::| |:: |:: :| | | :: | ::          | :::
gi-19224139.       MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP------FSIAL
                          10         20         30         40

                    70         80         90        100        110        120
gi-13621430. EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
             |::|  : : :    :: |:||:||  |||: |||| |:::|  ::|::|: | ||::|:
gi-19224139. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
             50         60         70         80         90        100

                   130        140        150        160        170        180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGI
             :|| |:|: |:| ::::| : |: ||| : || :   |   |:||:       |:
gi-19224139. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNI-----PKTP----
               110        120        130        140        150

                   190        200        210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
             || :||: | :: |:|||:: ::|  : ||||::
gi-19224139. LPLAGEVKSLLGILSIVLLGLLVLLYV-KKLKSKL
               160        170        180
```

FIGURE 59C

PCT/US05/27239

```
gi-13621430.pep
/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
generated symbols 1 to: 224.
 GETSEQ from morama, September 13, 2004 17:09.


SCORES   Init1: 163   Initn: 235   Opt: 264   z-score: 81.0  E(): 0.58
>>/home/morama/gas/pili/align/orf82.pep                    (224 aa)
 initn: 235 init1: 163 opt: 264 Z-score: 81.0 expect(): 0.58
Smith-Waterman score: 304;    30.5% identity in 213 aa overlap
 (1-213:30-222)

                                          10        20        30
gi-13621430.                              MKKSILRILAIGYLLMSFCLLDSVEAENLTA
                                          ::|    :::::.::|  : : ::| |:: |:
orf82.pep    LLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILTMLAFNQTVLAKDSTV
                    10        20        30        40        50        60

                40        50        60        70        80        90
gi-13621430. SINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVTTSVKGNGKTSFEQLTFS
             : :| | |.:: | ::       |  ::|::|  : : : :|  : |:||:||  |||:
orf82.pep    QTSISVENVLERAGDSTP------FSVALESIDAMKTI-DEIT--IAGSGKASFSPLTFT
                    70        80        90        100       110

              100       110       120       130       140       150
gi-13621430. EVGQYHYKIHQLLGKNSQYHYDETVYEVVIYVLYNEQSGALETNLVSNKLGETEKSELIF
              |||| |:::| ::|::|: | ||::|::|| |:|: |:| :::::| : |: ||| : |
orf82.pep    TVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDED-GTLVAKVISRRAGDEEKSAITF
                    120       130       140       150       160       170

              160       170       180       190       200       210
gi-13621430. KQEYSEKTPEPHQPDTTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATITLYSIYKKL
             | :   |   |:||:       |:      || :||: | :: |:|||:: ::|  : |||
orf82.pep    KPKRLVKPIPPRQPNI-----PKTP----LPLAGEVKSLLGILSIVLLGLLVLLYV-KKL
                    180       190       200       210       220

gi-13621430. KTSK
             |:
orf82.pep    KSRL


gi-13621430.pep
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS31
5]


SCORES   Init1: 164   Initn: 239   Opt: 261   z-score: 80.5  E(): 0.62
>>/home/morama/gas/pili/align/gi-21909638.pep              (183 aa)
 initn: 239 init1: 164 opt: 261 Z-score: 80.5 expect(): 0.62
Smith-Waterman score: 302;    31.5% identity in 200 aa overlap
 (14-213:2-181)

                    10        20        30        40        50        60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
                      :|  : : ::| |:: |:: :| | | :: | ::         | :::
gi-21909638.           MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP------FSIAL
```

FIGURE 59D

PCT/US05/27239

```
                                    10        20        30        40

                    70        80        90       100       110       120
gi-13621430. EALDKESPLPNSVTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
             |::|  : : :    :: |:||:||  |||: |||| |:::|  ::|::|: | ||::|:
gi-21909638. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
                   50        60        70        80        90

                  130       140       150       160       170       180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGI
             :|| |:|: |:| :::::| : |: ||| : || ::  |   |:||:      |:
gi-21909638. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNI-----PKTP----
           100       110       120       130       140

                  190       200       210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
             || :||: | :: |:|||:: ::|  : ||||:
gi-21909638. LPLAGEVKSLLGILSIVLLGLLVLLYV-KKLKSRL
           150       160       170       180
```

gi-13621430.pep
/home/morama/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP_606439.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 121 Initn: 121 Opt: 126 z-score: 61.4 E(): 6.7
>>/home/morama/gas/pili/align/gi-19745303.pep (344 aa)
initn: 121 init1: 121 opt: 126 Z-score: 61.4 expect(): 6.7
Smith-Waterman score: 126; 27.0% identity in 100 aa overlap
(59-155:84-183)

```
             30        40        50        60        70        80
gi-13621430. LTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLP--NSVTTSVKGNG-KTSF
                                              ||::|:: : :   ||  |::| :: : :|
gi-19745303. LMPKADYTFKVEADDNAKGKTKDGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDF
                   60        70        80        90       100       110

                   90       100       110       120       130       140
gi-13621430. EQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVVIYVLYNEQSGALETNLVSNKLGETE
              :: |  || |:| : :: |:::   ||  : | :||:  |::|    .:||:: |:::
gi-19745303. ANVKFPGVGVYRYTVSEVNGNKAGIAYDSQQWTVDVYVVNREDGGFEAKYIVSTEGGQSD
                  120       130       140       150       160       170

                  150       160       170       180       190       200
gi-13621430. KSELIFKQEYSEKTPEPHQPDTTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATITLY
             |: ::||: ::
gi-19745303. KKPVLFKNFFDTTSLKVTKKVTGNTGEHQRSFSFTLLLTPNECFEKGQVVNILQGGETKK
                  180       190       200       210       220       230
```

gi-13621430.pep
/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 58 Initn: 86 Opt: 122 z-score: 60.9 E(): 7.2
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
initn: 86 init1: 58 opt: 122 Z-score: 60.9 expect(): 7.2
Smith-Waterman score: 135; 29.1% identity in 172 aa overlap

FIGURE 59E

```
                    10        20        30        40        50
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVD-VATNKQSSDIDETFMF-
                    :|  |   | ||    :|::|::: : :: | :::| |':|    : | || :
gi-13621428.   MKLRHLLLLTGAALTSFAAT-TVHGETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIE
                      10        20        30        40        50

                   60                70        80        90              100
gi-13621430. ----VIEALDK------ESPLPN-SVTTSVKGNGKTSFEQLTFSEV-----GQYHYKI-H
                  | |  :|     :|:.: : |:| ||:::|:  :: ||||     | |:||: :
gi-13621428. PDTTVNEDGNKFKGVALNTPMTKVTYTNSDKGGSNTKTAEFDFSEVTFEKPGVYYYKVTE
                60        70        80        90        100       110

                    110       120       130       140       150       160
gi-13621430. QLLGKNSQYHYDETVYEVVIYVLYNE-QSGALETNLVSNKLGETEKSELIFKQEYSEKTP
                : :.|     || | |.| ::||:|| |:  : | :|: | |   |  : ||:  :  |
gi-13621428. EKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVATYIVGYKEGS--KVPIQFKNSLDSTTL
                120       130       140       150       160       170

                    170       180       190       200       210
gi-13621430. EPHQPDTTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK

gi-13621428. TVKKKVSGTGGDRSKDFNFGLTLKANQYYKASEKVMIEKTTKGGQAPVQTEASIDQLYHF
                  180       190       200       210       220       230


gi-13621430.pep
/home/morama/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]

SCORES   Init1: 88    Initn: 88    Opt: 119   z-score: 60.4  E(): 7.5
>>/home/morama/gas/pili/align/gi-19224137.pep            (342 aa)
 initn:  88 init1:  88 opt: 119 Z-score: 60.4 expect():  7.5
Smith-Waterman score: 119;    29.7% identity in 111 aa overlap
(72-176:93-201)

                    50        60        70        80              90
gi-13621430. DVATNKQSSDIDETFMFVIEALDKESPLPNSVTTSVKGNGK-----TSFEQLTFSEVGQY
                                          | :::. | :||     ::| ::||  || |
gi-19224137. SVNPDSAATGTESNLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIY
                   70        80        90       100       110       120

                  100       110       120       130       140       150
gi-13621430. HYKIHQLLGKNSQYHYDETVYEVVIYVLYNEQSGALETNLVSNKLGETEKSELI-FKQEY
                :| :  :   |     ||:| : | :||  ||:: | || : : :| |::  :| | |.:: :
gi-19224137. RYVVTENKGTAEGVTYDDTKWLVDVYVGNNEKGG-LEPKYIVSKKGDSATKEPIQFNNSF
                   130       140       150       160       170       180

                  160       170       180       190       200       210
gi-13621430. SEKTPEPHQPDTTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
                | |    : ::| :   :||
gi-19224137. -ETTSLKIEKEVTGNTGDHKKAFTFTLTLQPNEYYEASSVVKIEENGQTKDVKIGEAYKF
                   190       200       210       220       230       240


gi-13621430.pep
/home/morama/gas/pili/align/gi-50913503.pep

gi|50913503|ref|YP_059475.1| Fibronectin-binding protein [Streptococcus pyogenes
 MGAS10394]
```

FIGURE 59F

PCT/US05/27239

```
SCORES   Init1: 73    Initn: 73    Opt: 117   z-score: 60.4  E(): 7.6
>>/home/morama/gas/pili/align/gi-50913503.pep              (627 aa)
 initn:  73 init1:  73 opt: 117 Z-score: 60.4 expect():  7.6
Smith-Waterman score: 118;    28.7% identity in 87 aa overlap
 (129-215:549-625)

            100       110       120       130       140       150
gi-13621430. KIHQLLGKNSQYHYDETVYEVVIYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEK
                                          ||  ||  |     |  : :|:|:. : :|
gi-50913503. IETEDTKEPEVLMGGQSESVEFTKDTQTGMSGFSETATVV----EDTRPKLVFHFDNNEP
            520       530       540       550             560       570

            160       170       180       190       200       210
gi-13621430. TPEPHQPDTTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
               | :      :||| |: : |||:||:: : :: |||::::::::::|: |: :::|
gi-50913503. KVEEN------REKPTKNITPILPATGDIENVLAFLGILILSVLSIFSLLKNKQSNKKV
                        580       590       600       610       620


gi-13621430.pep
/home/morama/gas/pili/align/gi-19224134.pep

gi|19224134|gb|AAL86405.1|AF447492_2 protein F [Streptococcus pyogenes]


SCORES   Init1: 73    Initn: 73    Opt: 115   z-score: 60.1  E(): 7.8
>>/home/morama/gas/pili/align/gi-19224134.pep              (698 aa)
 initn:  73 init1:  73 opt: 115 Z-score: 60.1 expect():  7.8
Smith-Waterman score: 115;    27.4% identity in 73 aa overlap
 (143-215:631-697)

                  120       130       140       150       160       170
gi-13621430. DETVYEVVIYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEPHQPDTTEKEK
                                          |  : :|:|: : :|   | :       :||
gi-19224134. VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDNNEPKVEEN------REK
                        610       620       630       640       650

                  180       190       200       210
gi-13621430. PQKKRNGILPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
               | |: : |||:||:: : :: |||::::::::::|: |: :::|
gi-19224134. PTKNITPILPATGDIENVLAFLGILILSVLSIFSLLKNKQNNKV
                  660       670       680       690


! Distributed over 1 thread.
!      Start time: Wed Sep 15 18:45:36 2004
! Completion time: Wed Sep 15 18:45:38 2004

! CPU time used:
!        Database scan:  0:00:00.1
! Post-scan processing:  0:00:00.3
!       Total CPU time:  0:00:00.4
! Output File: gi-13621430.fasta
```

FIGURE 59G

PCT/US05/27239

TYPE 3 pilus motifs

protein F2 like fibronectin-binding protein
Length: 696-733
LPXTG
pilin motif consensus PK (X_7) K
E box consensus ETxAPxGY

SpyM3_0104/21909640	pilin motif	155 PKEKPIIYFK
	E box	398 YTFVETAAPDGY
SPs0106/28895018	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
SpyM18_0132/19745307	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
orf84	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY

TYPE 4 pilus motifs

protein F2 like fibronectin-binding protein
Length: 1161
LPXTG
pilin motif consensus PK (X_{7-8}) K
E box consensus YxLxETxAPxGY

The protein is longer than the proteins belonging to TYPE 3 and has 4 possible pilin motifs and 2 E boxes

19224141	pilin motifs	215 PKGISQDIPK 571 PKGYQQVTEK 156 PKMSVVSKYGK 674 PKYDAKNQEYK
	E boxes	563 YDLYETKAPKGY 940 YTFVETAAPDGY

FIGURE 60

Formation of pili structures on GBS appears to be correlated to FACS values for surface expression of GBS80 protein

Immunogold Electron Microscopy

α-GBS80

α GBS80

COH1
Δmean=300

JM9130013
Δmean=500

CHIRON | VACCINES

FIGURE 61

Surface exposure is capsule-dependent for GBS 322 but not for GBS 80

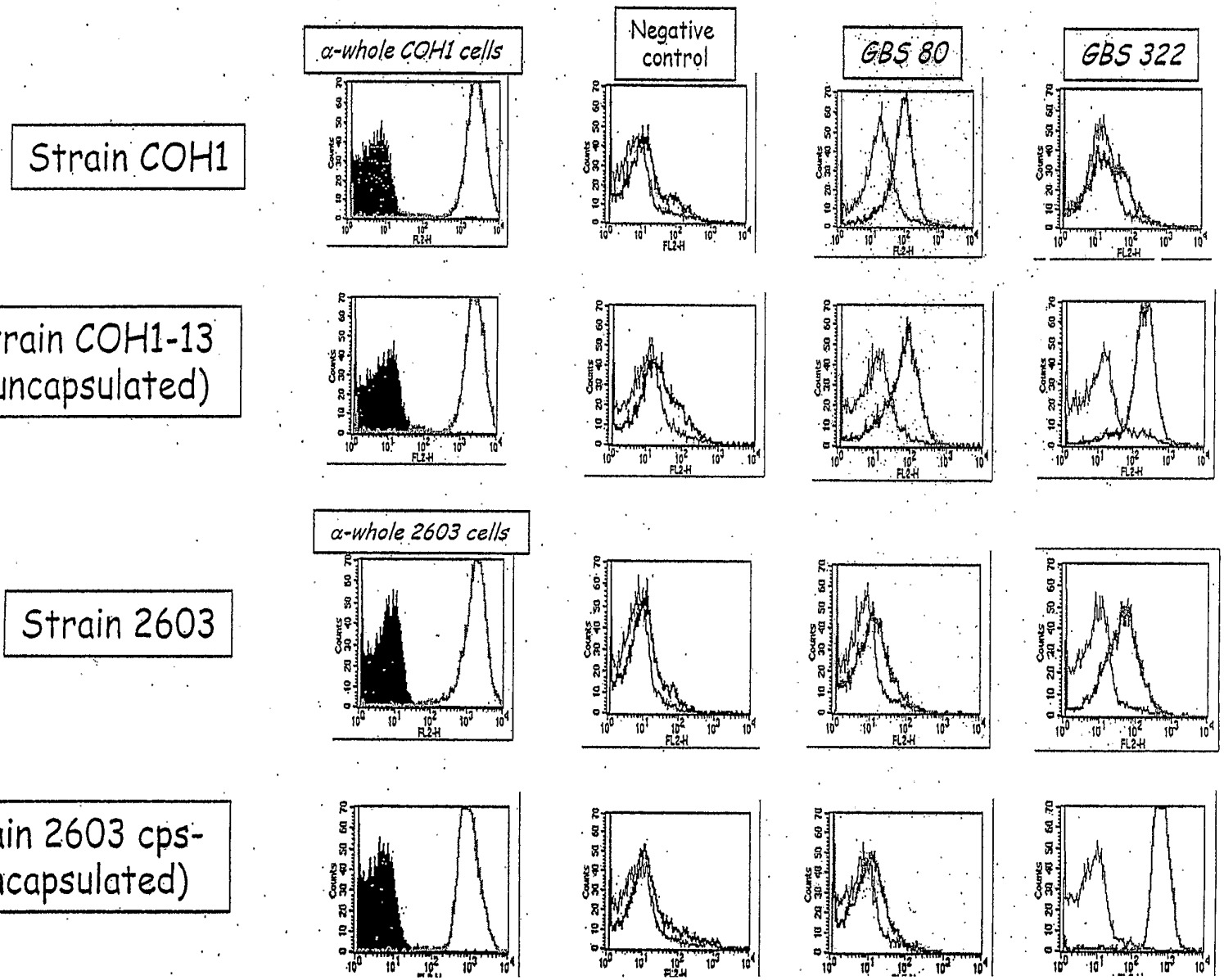
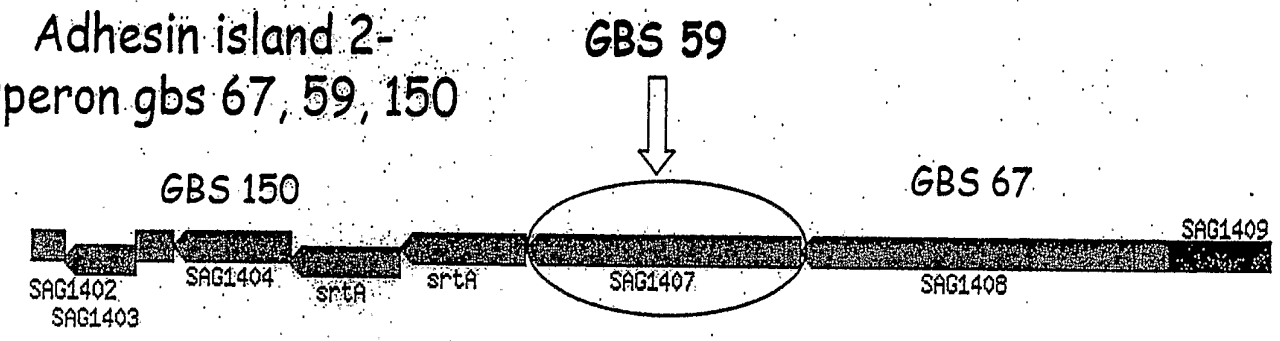


FIGURE 62

Adhesin island 2- Operon gbs 67, 59, 150



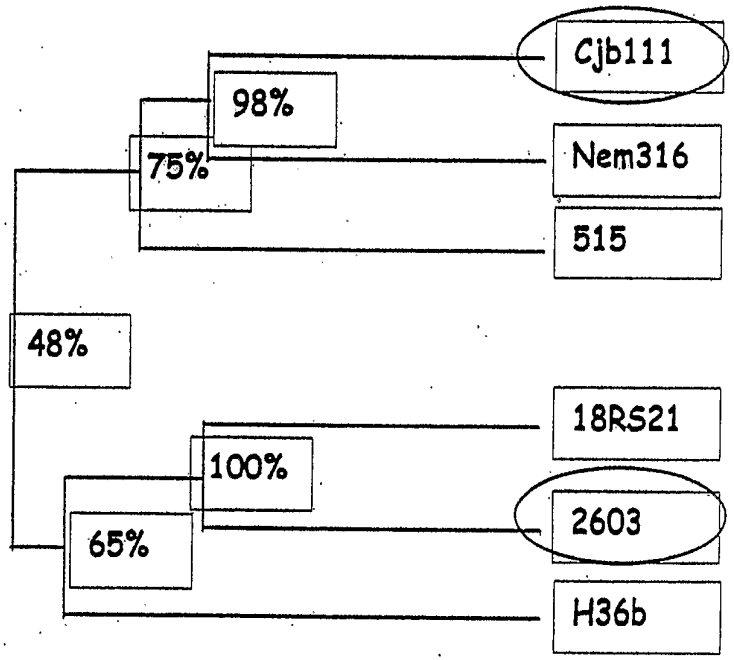
- 30 % identity with GBS 80

- By comparing GBS 59 amino acidic sequence of 2603 with that of other sequenced genomes, the following homologies are obtained:

2603	
nem316	47%
h36b	62%
515	48%
cjb111	48%
18rs21	100%
coh1	not present (Spb1)
A909	not present (Spb1)

- CGH: 1/20 GBS strains analyzed (18RS21)

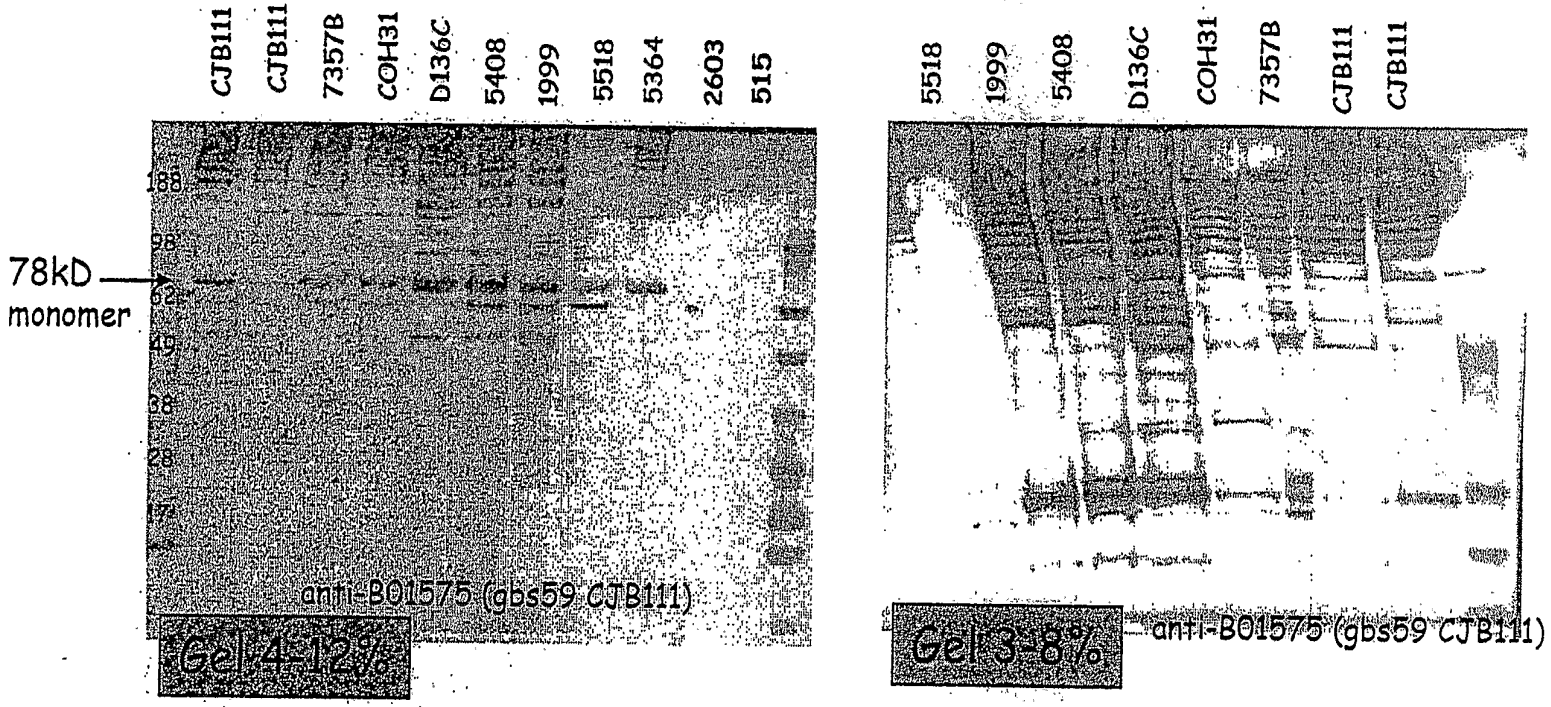
Two-by-two amino acidic sequence comparison



There seem to be two clusters suggesting the presence of two major isoforms

FIGURE 63

Western blotting with whole extracts derived from GBS strains



GBS 59 is part of a high molecular weight polymer (pilus)

in GBS strains: CJB111, 7357b, coh31, d1363c, 5408, 1999, 5364, 5518, 515

FIGURE 64

Western blotting with purified proteins and whole extracts derived from GBS strains

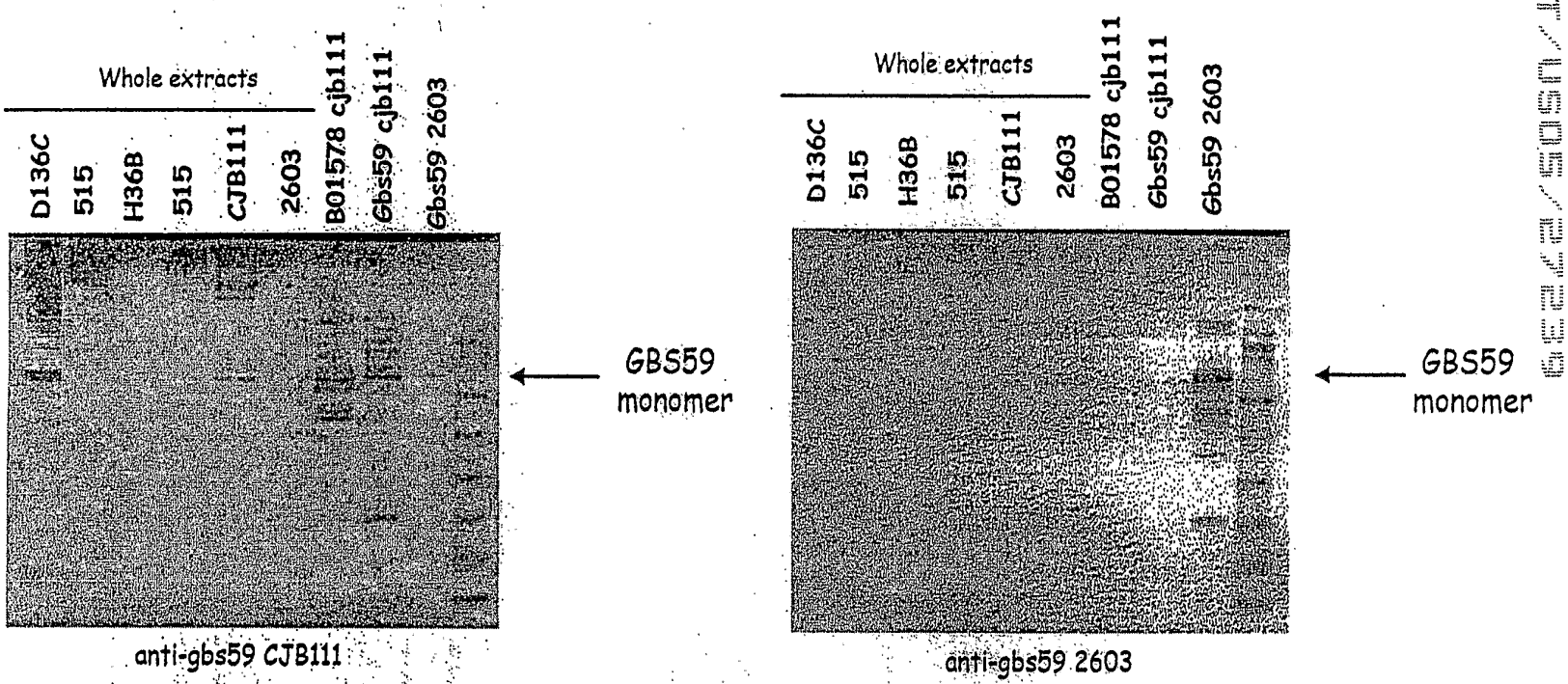
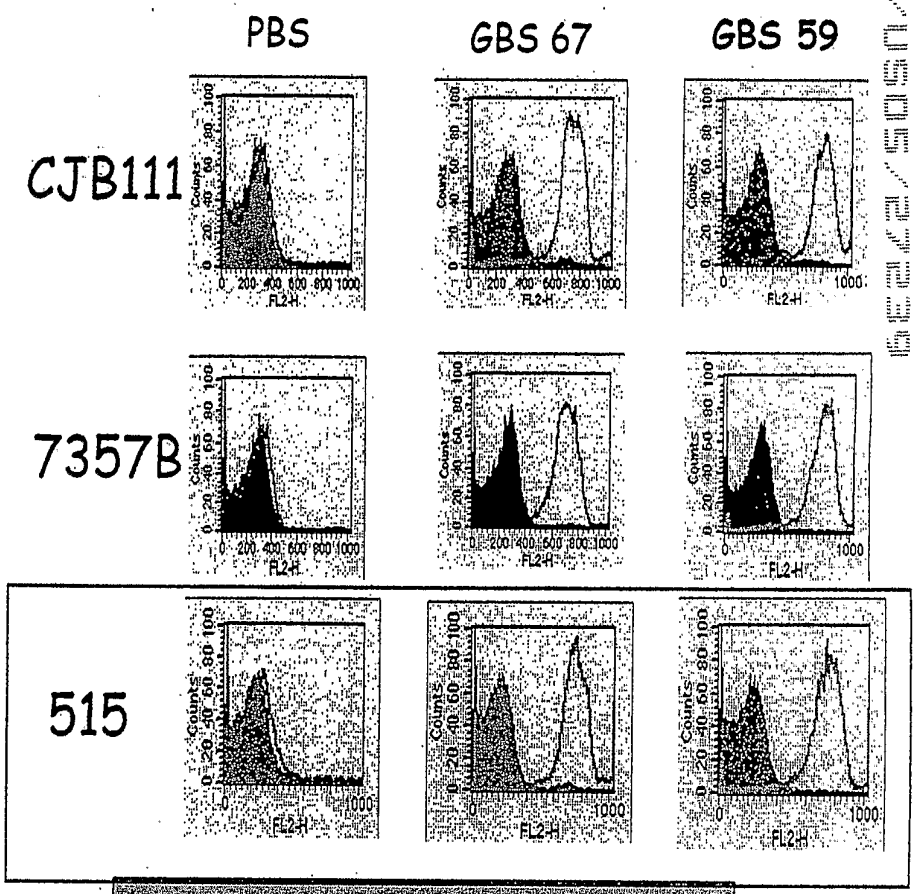


FIGURE 65

FACS analysis using mouse antiserum after immunization with BO1575 (gbs59) from CJB111 genome

		GBS 59
GBS strains	Type	FACS (Δ Mean)
DK1	Ia	565
DK8		559
Davis		577
515		583
2986		443
5551		524
7357b-		596
5518		190
D136C	III	504
COH31		505
DK21	II	249
CJB111	V	493
5364		593
2110		590
1999		594
2210		636
5408		537
1169		227

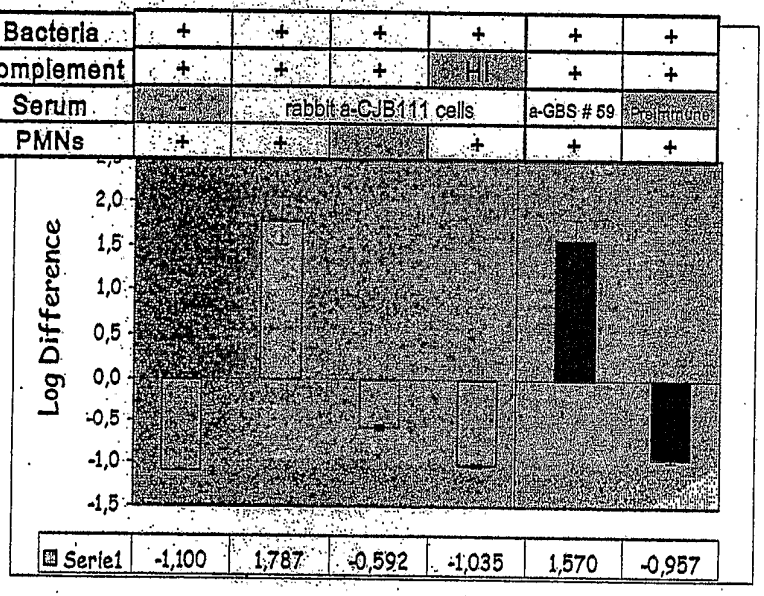


Where present GBS 59 is a highly exposed on the GBS surface

FIGURE 66

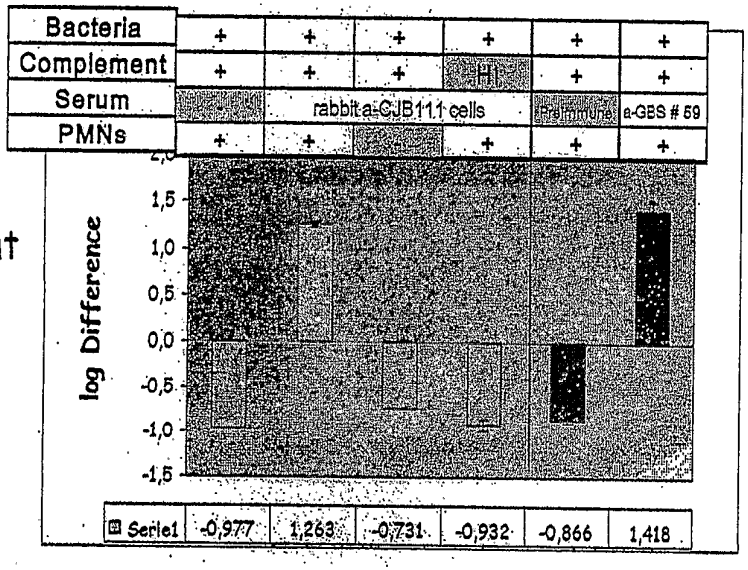
Opsonophagocytosis assays: BO1575 (gbs59-cjb111)

I experiment



- CJB111 GBS strain type Ia
- Baby rabbit complement
- Human PMNs
- Positive control: anti-type V cells (rabbit serum anti fixed type V cells)

II experiment



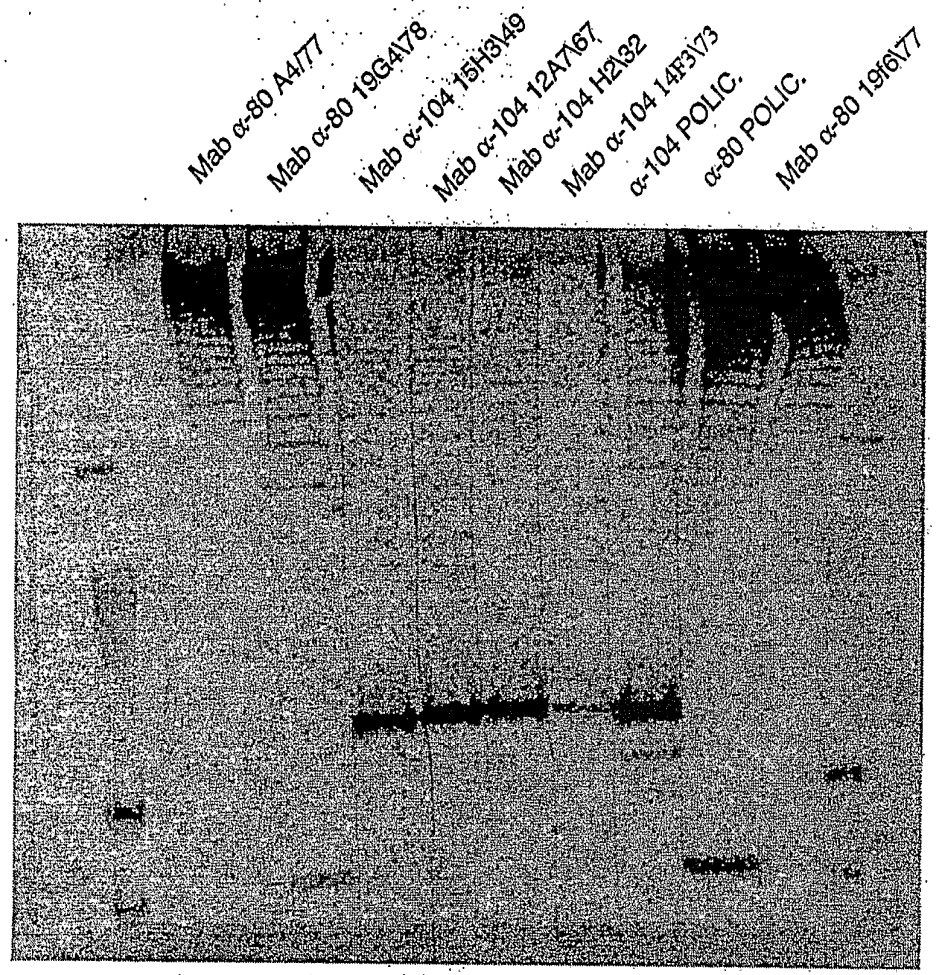
Antibodies against BO1575 (cjb111) are opsonic for cjb111 GBS strain serotype V

FIGURE 67

WO 2006/078318 PCT/US2005/027239 217/487 PCT/US2005/027239

Association GBS 80-104

WB α-80, α-104 JM9130013 Total Extract

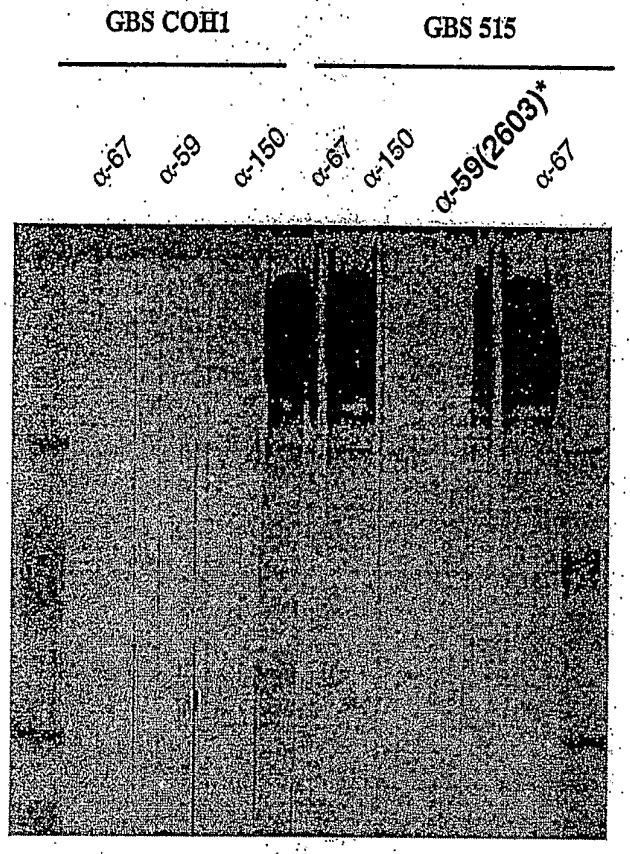


FACS	(ΔMEAN)
GBS 80	597
GBS 104	446

FIGURE 68

WO 2006/078318 PCT/US2005/027239
218/487
PCT/US2005/027239

WB GBS 515 Total Extract
α-67; α-150; α-59 (2603)



Controls:
GBS COH1 total extracts

- anti-gbs59 mouse serum after immunization with SAG1407 (GBS 59) from 2603 genome

GBS 67 and GBS 150 are parts of
a high molecular weight polymer (pilus)
in 515 GBS strain

FIGURE 69

WO 2006/078318 PCT/US2005/027239 219/487 PCT/US2005/027239

PCT/US05/27239

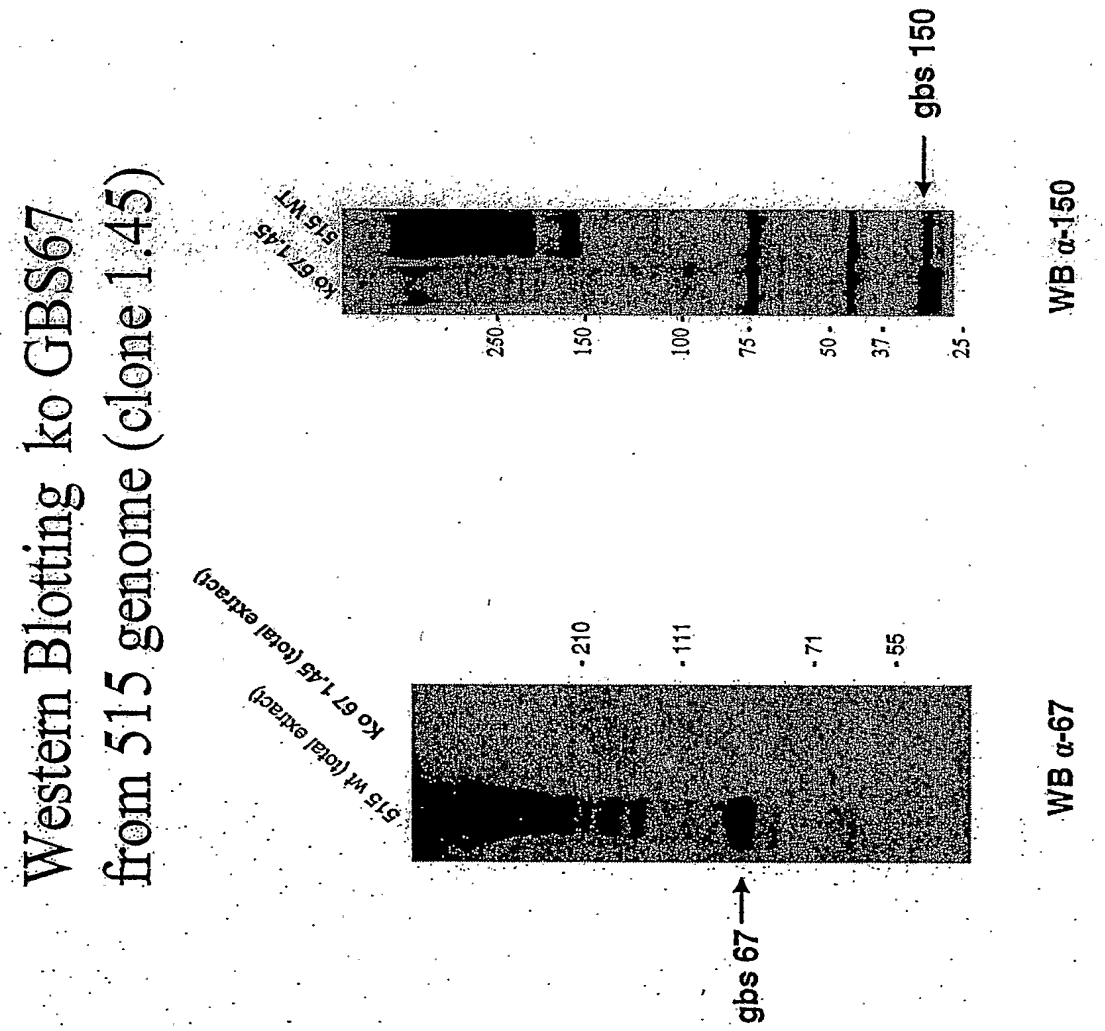


FIGURE 70

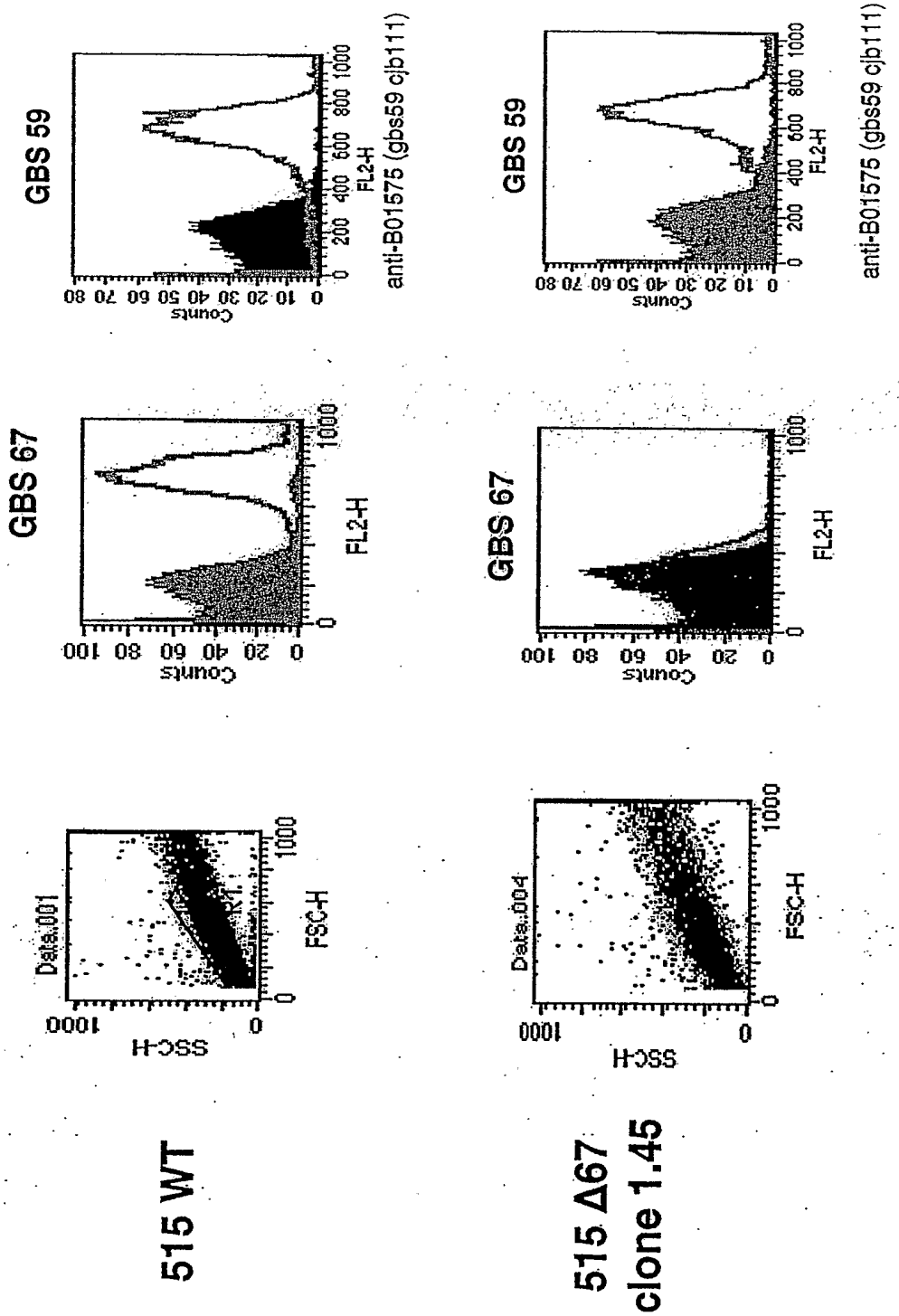
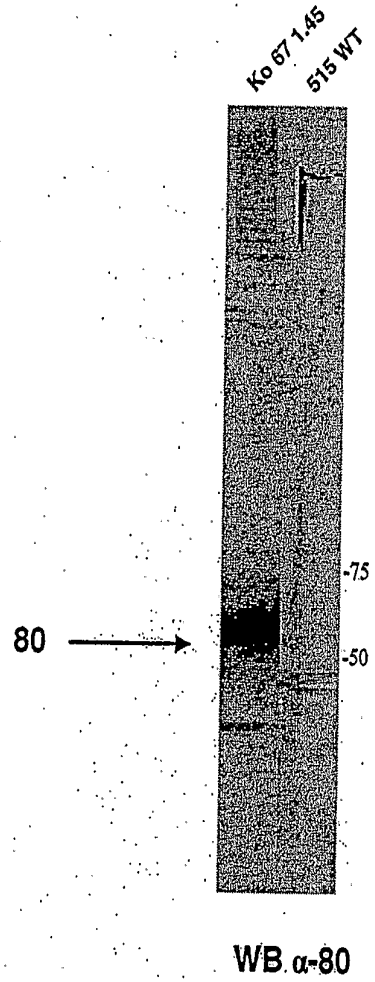

FACS GBS 515 Δ67
GBS 59
GBS 67
anti-B01575 (gbs59 c|b111)
Counts
FL2-H
SSC-H
FSC-H
Data.001
Data.004
515 WT
515 Δ67
clone 1.45

FIGURE 71

Complementation of GBS 515 KO 67 with pAM401-gbs80



GBS 80 forms a high molecular weight complex (pilus) in absence of GBS 67

FIGURE 72

spyM6_0159 type 1 pilus present in M6

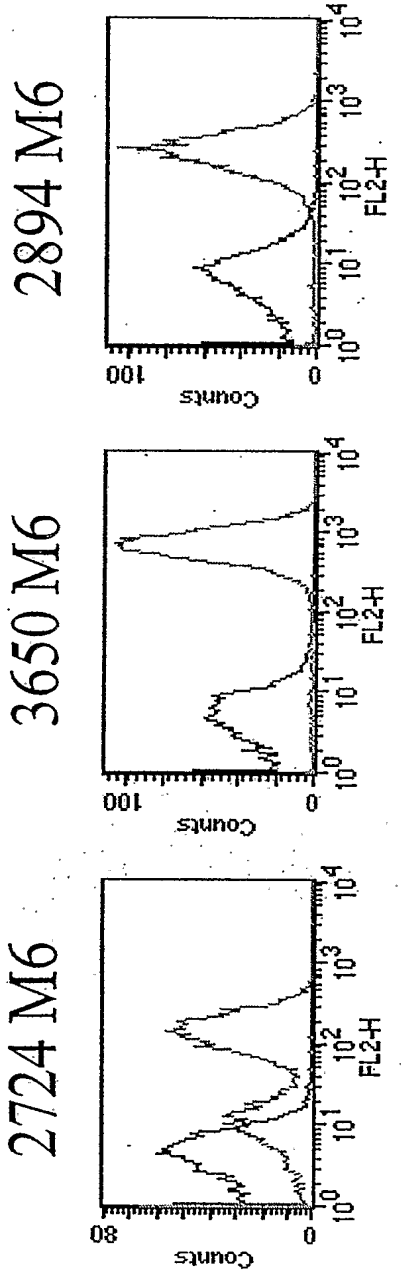


Figure 73

spyM6_0160 type 1 pilus present in M6

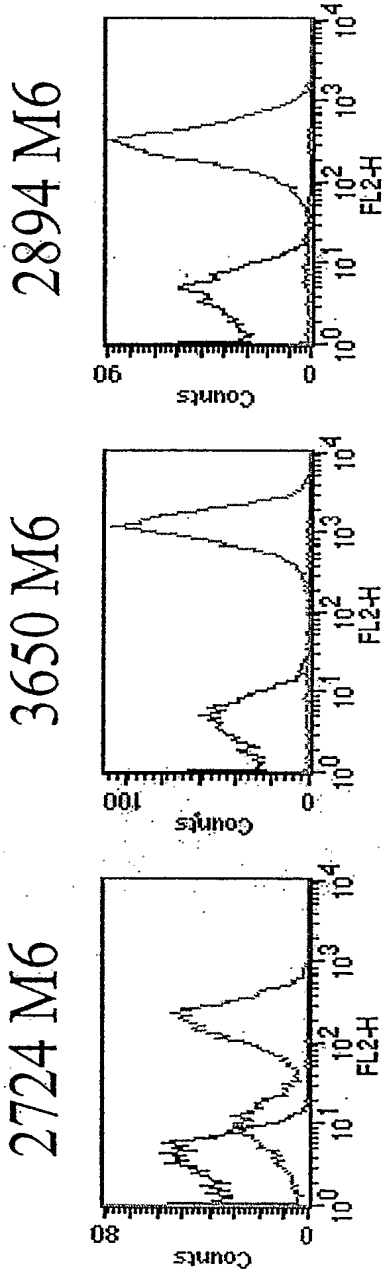


Figure 74

PCT/US05/27239

Gas15 type 2 pilus present in M1

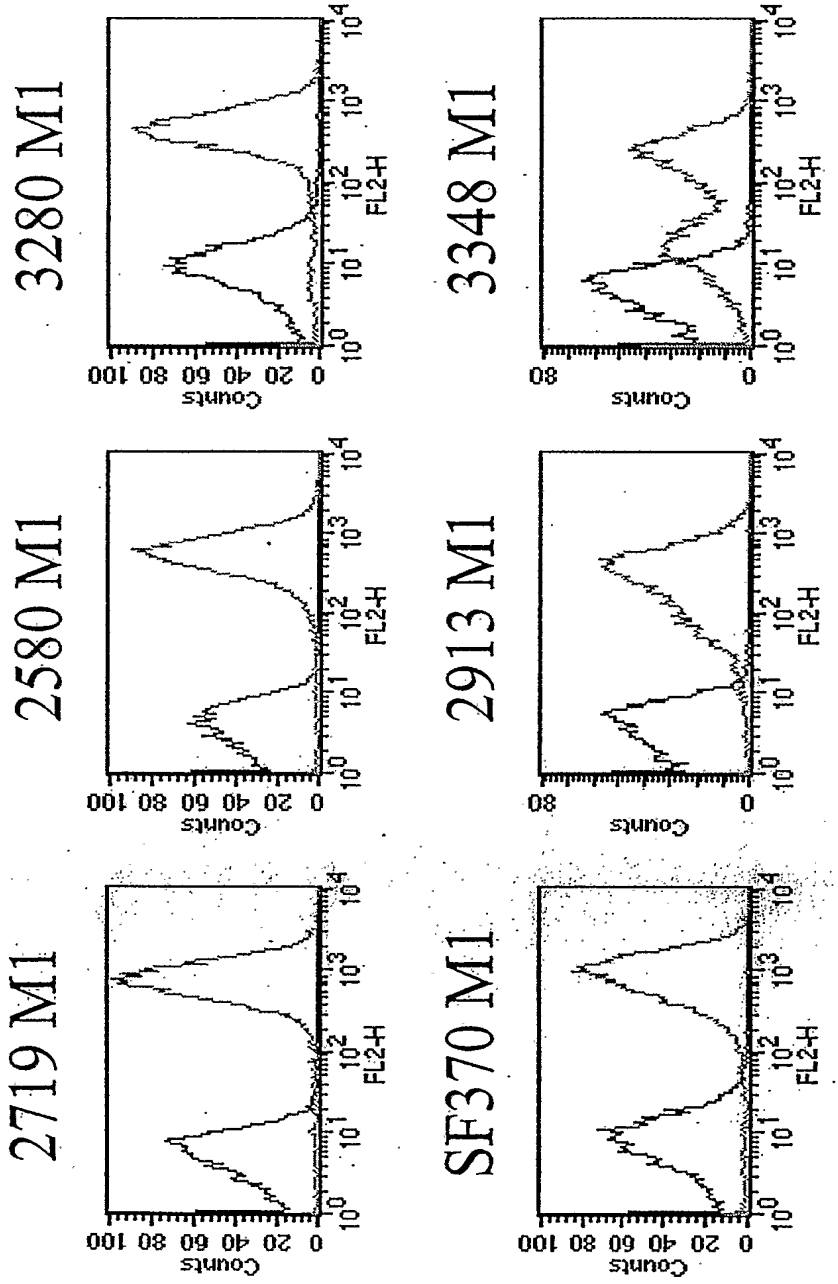


Figure 75

PCT/US05/27239

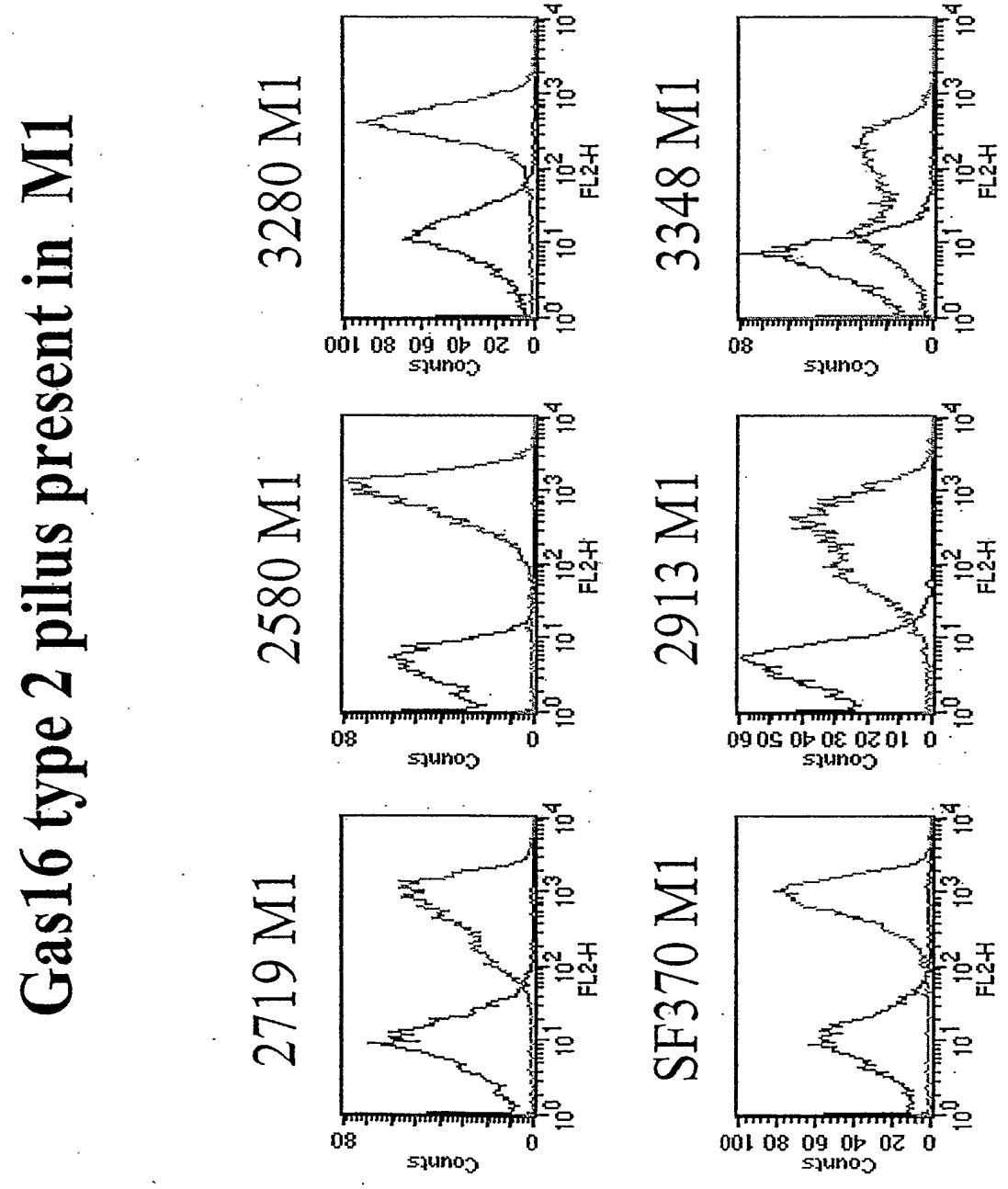


Figure 76

PCT/US05/27239

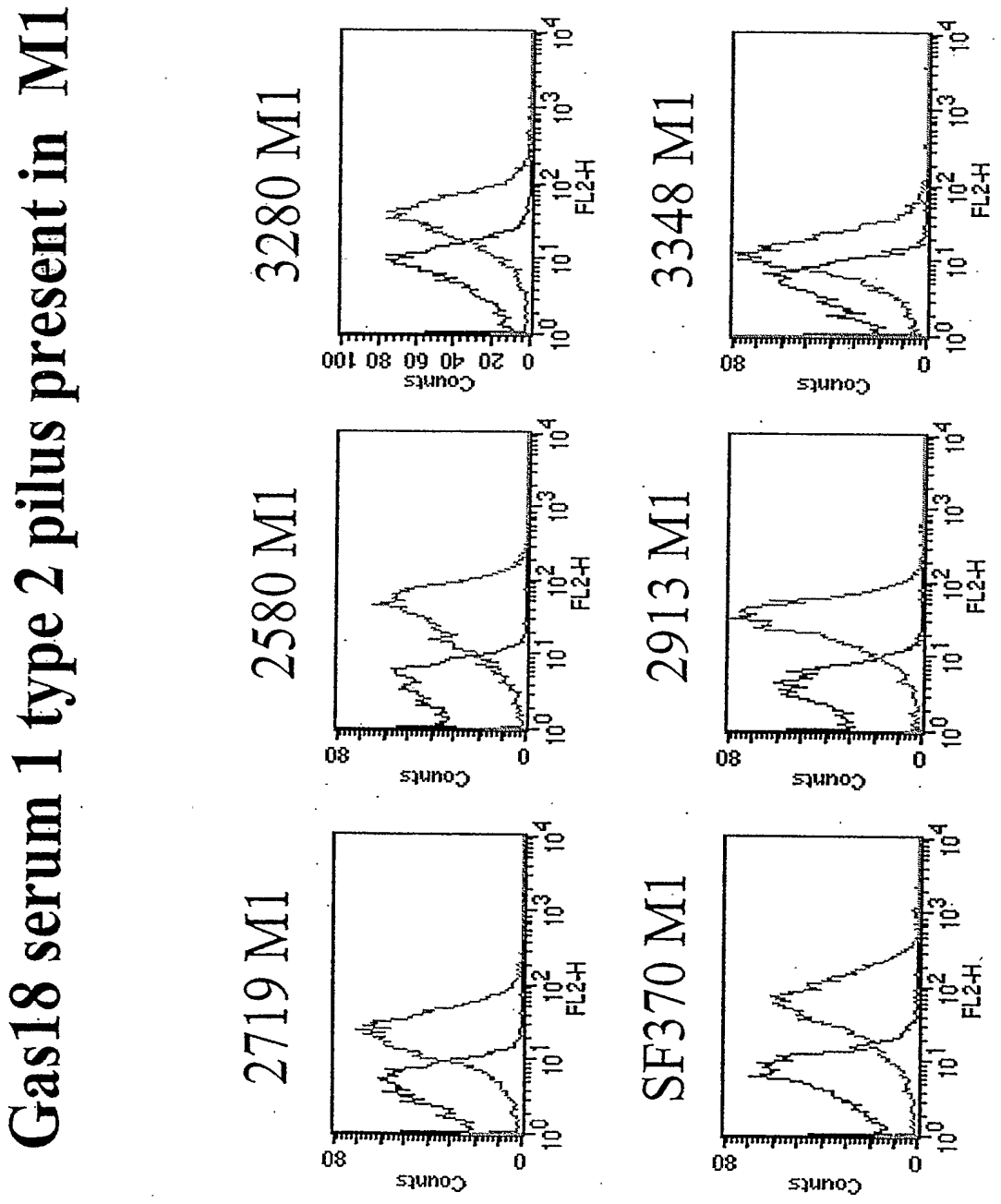


Figure 77

PCT/US05/27239

228/487

Gas18 serum 2 type 2 pilus present in M1

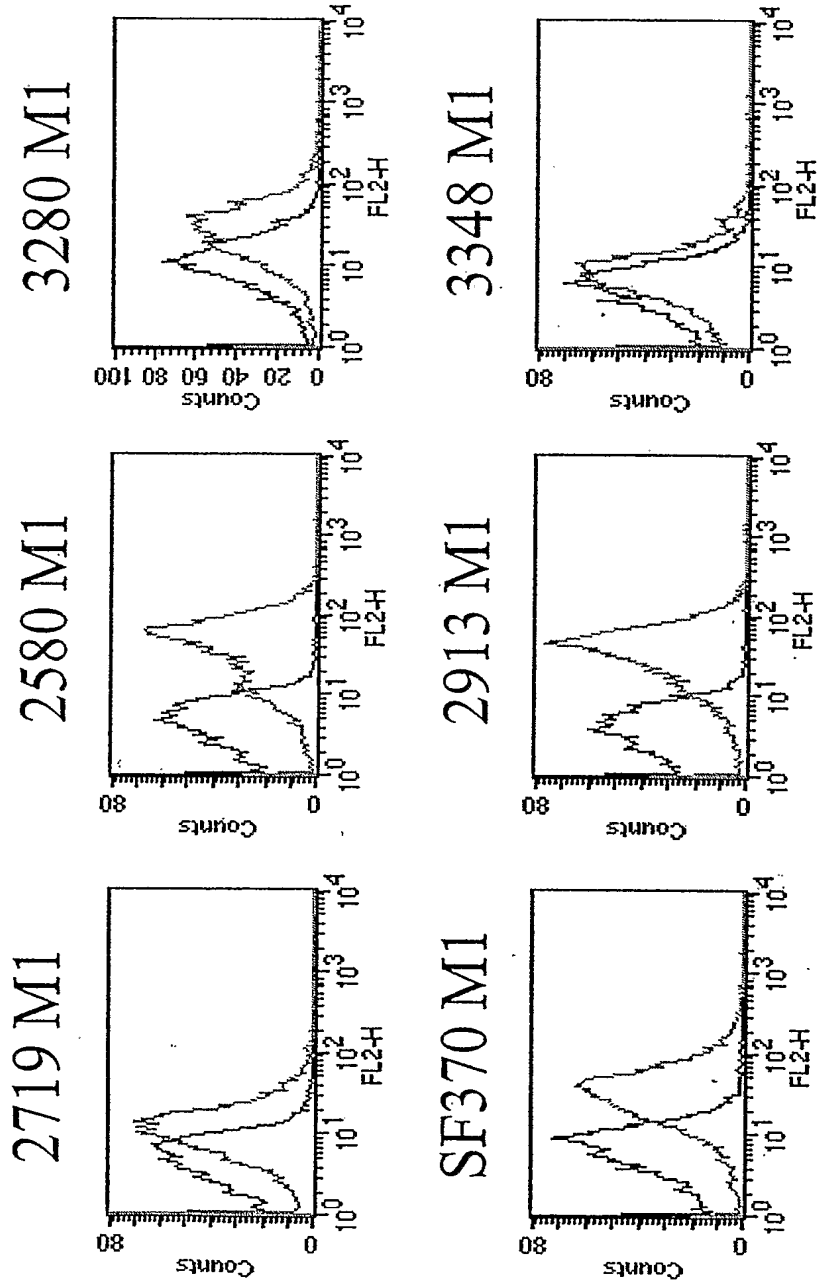


Figure 78

PCT/US05/27239

Gas16p2 type 2 pilus present in M1

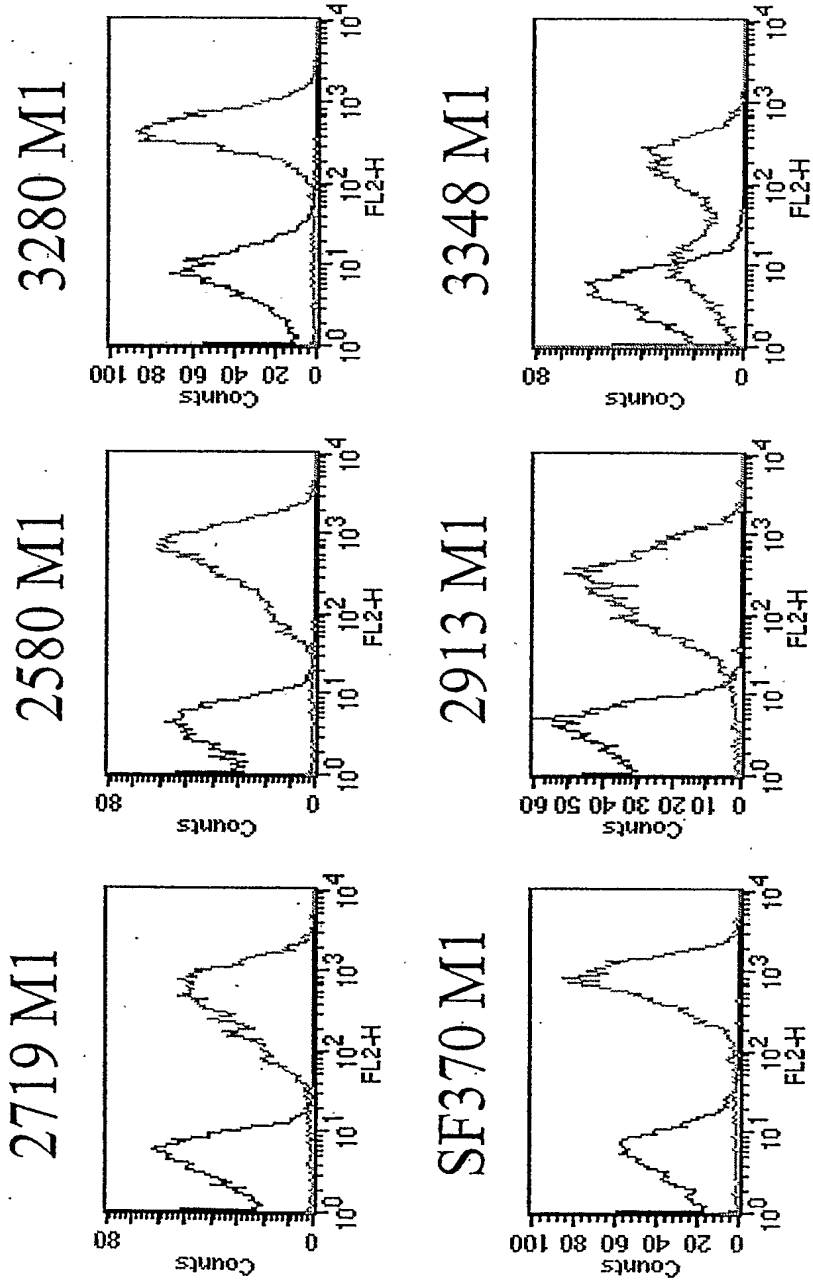


Figure 79

PCT/US05/27239

spyM3_0098 type 3 pilus present in M3

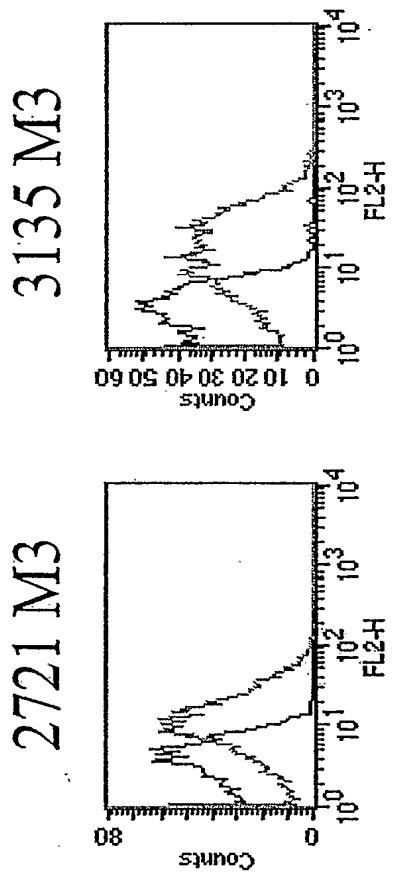


Figure 80

PCT/US05/27239

spyM3_0100 type 3 pilus present in M3

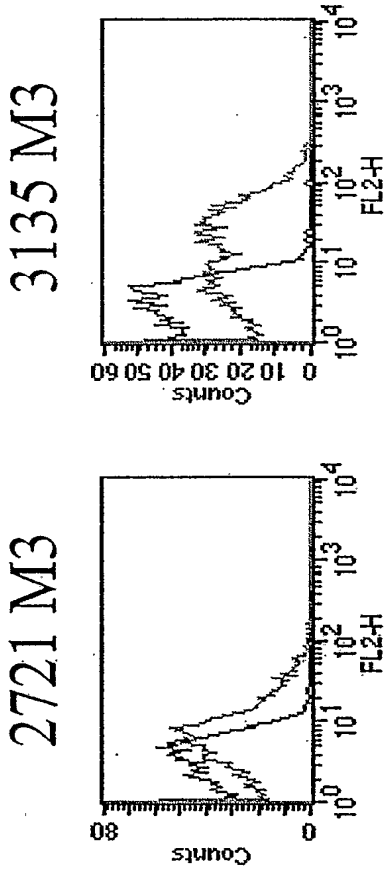


Figure 81

spyM3_0102 type 3 pilus present in M3

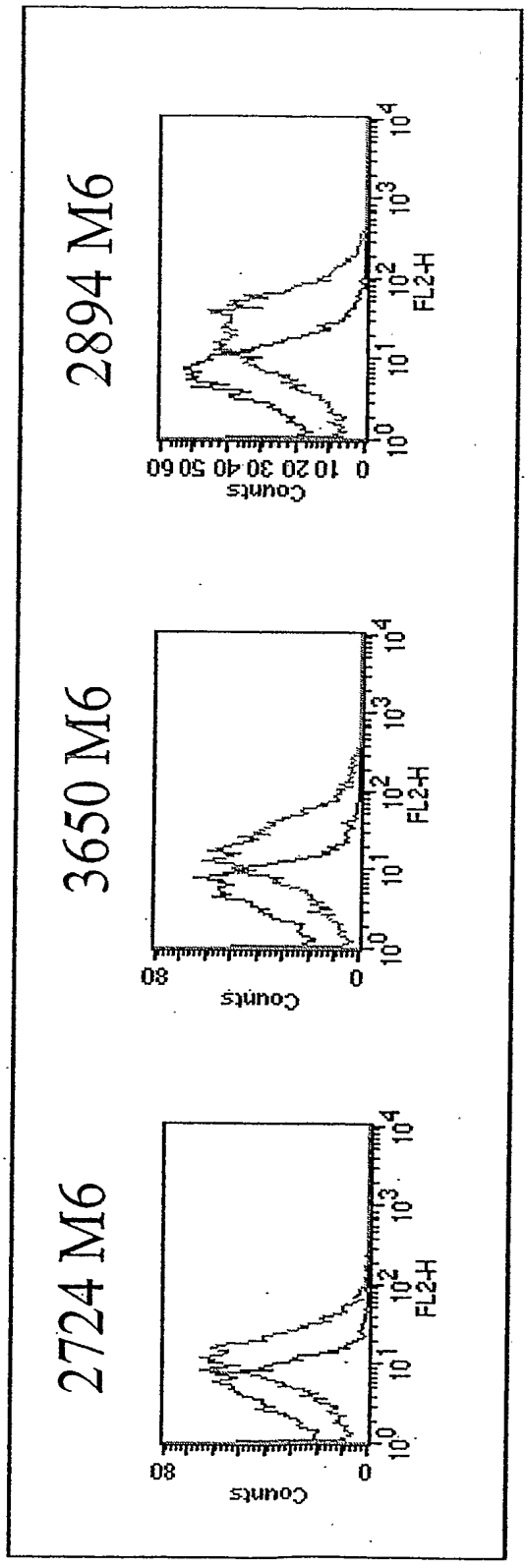


Figure 82

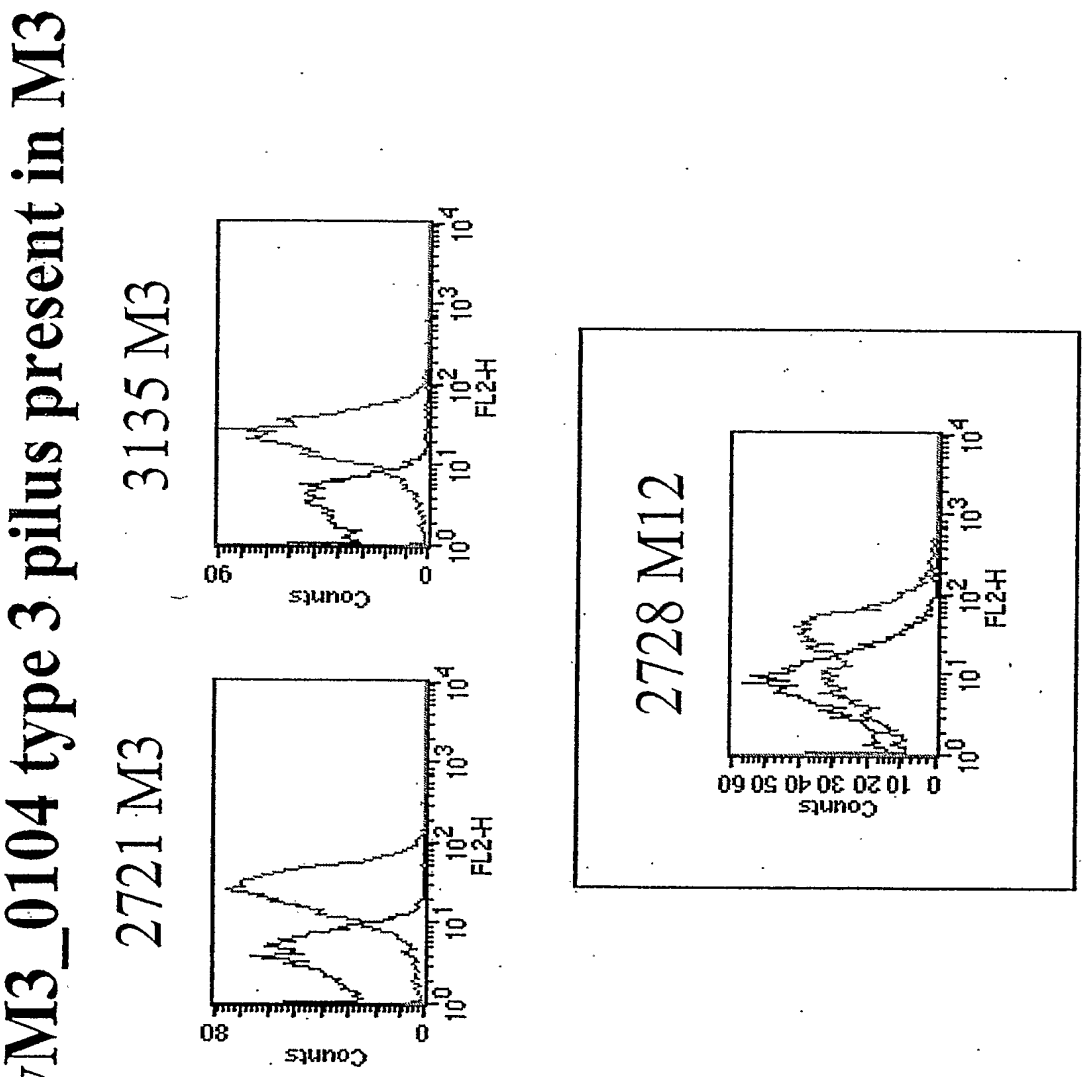

spyM3_0104 type 3 pilus present in M3
2721 M3
3135 M3
2728 M12
Counts
FL2-H

Figure 83

spyM3_0106 type 3 pilus present in M3

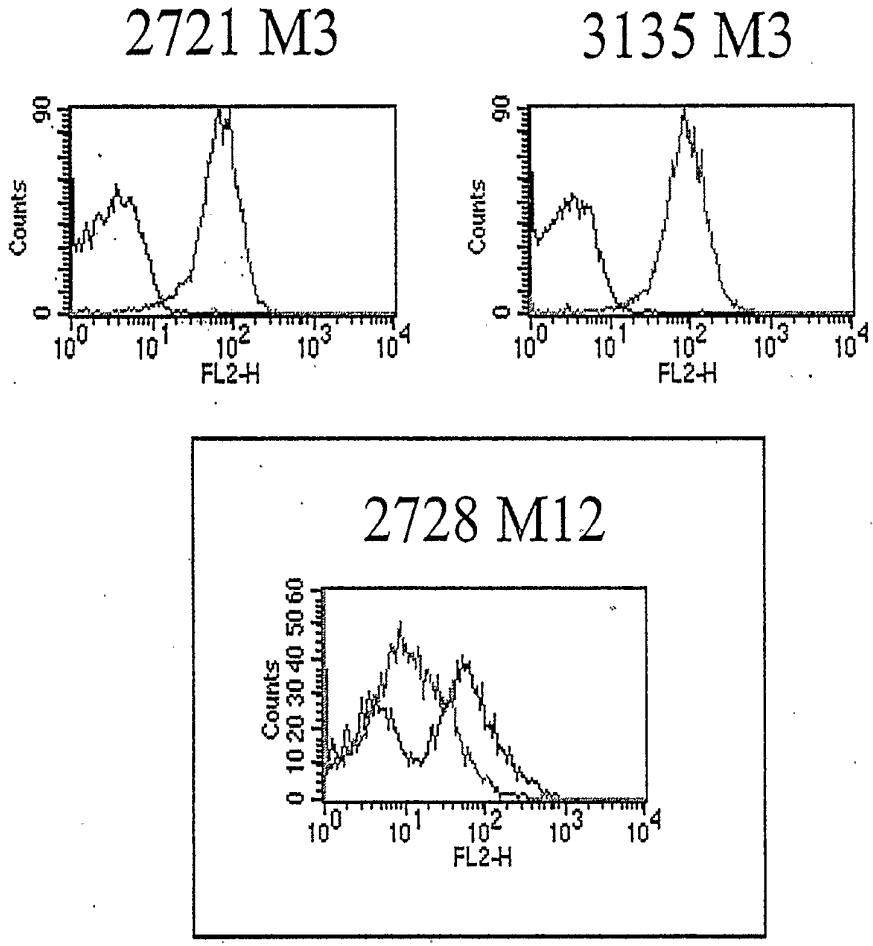


Figure 84

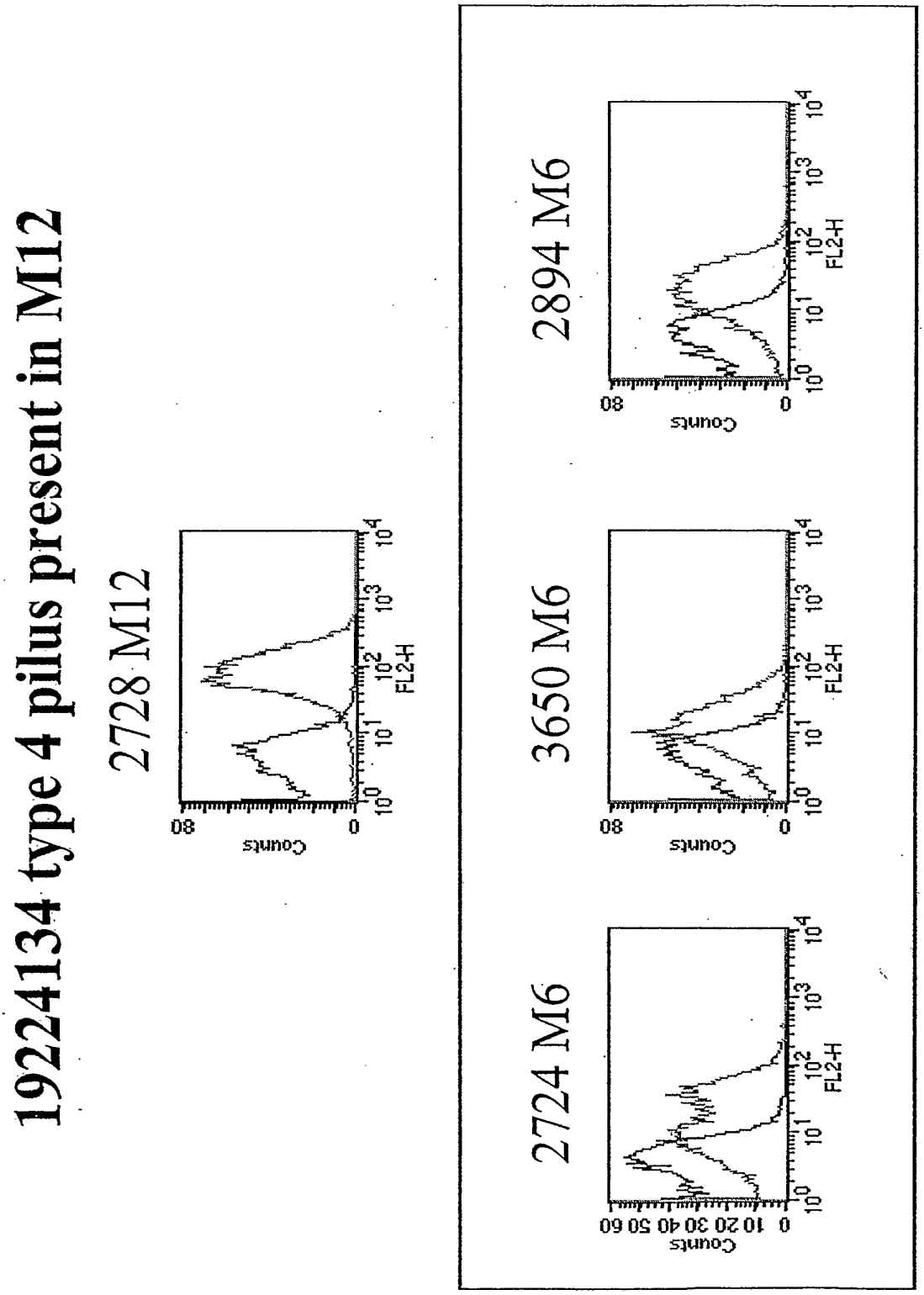

19224134 type 4 pilus present in M12
2728 M12
2724 M6
3650 M6
2894 M6
Counts
FL2-H
0
80
0 10 20 30 40 50 60

Figure 85

19224135 type 4 pilus present in M12

2728 M12

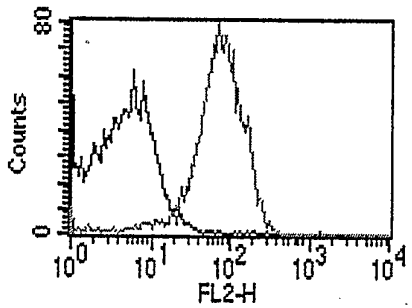


Figure 86

19224137 type 4 pilus present in M12

2728 M12

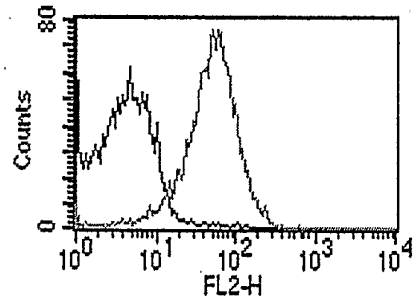


Figure 87

19224141 type 4 pilus present in M12

2728 M12

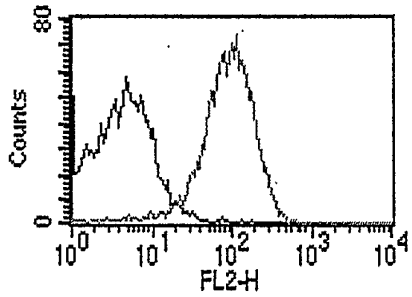
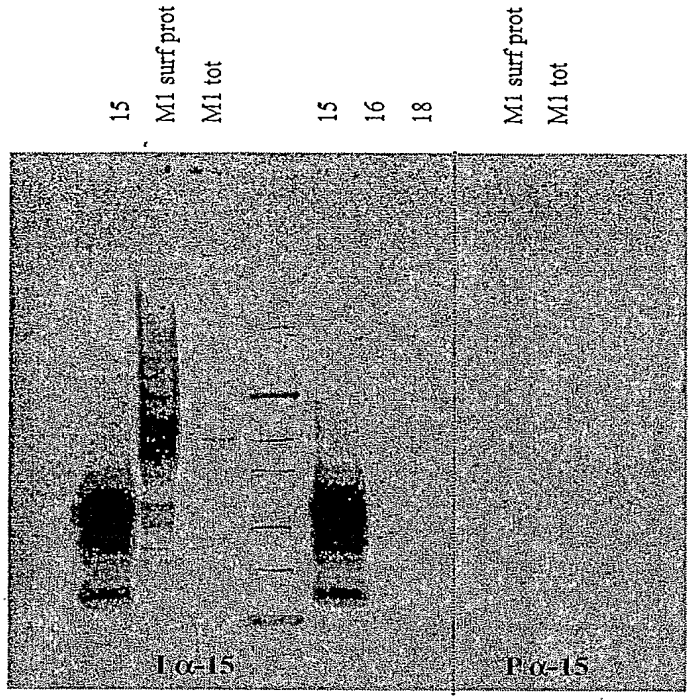


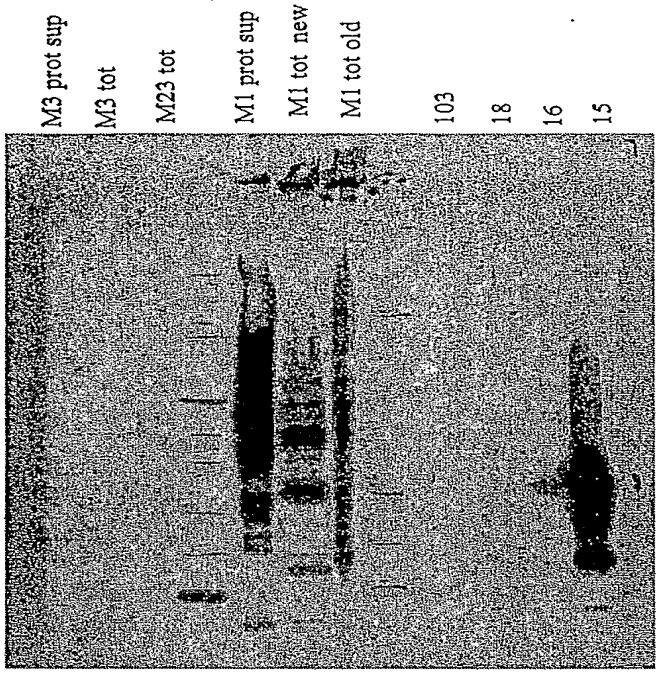
Figure 88



LEGEND:

M1 tot: total extract (M1)
M1 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 89



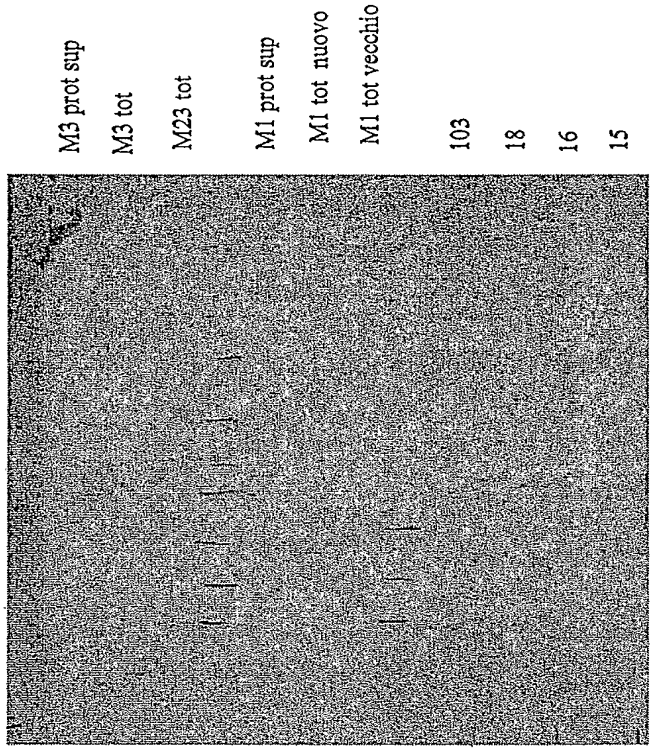
I α-15

LEGENDA:

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 90

PCT/US2005/027239



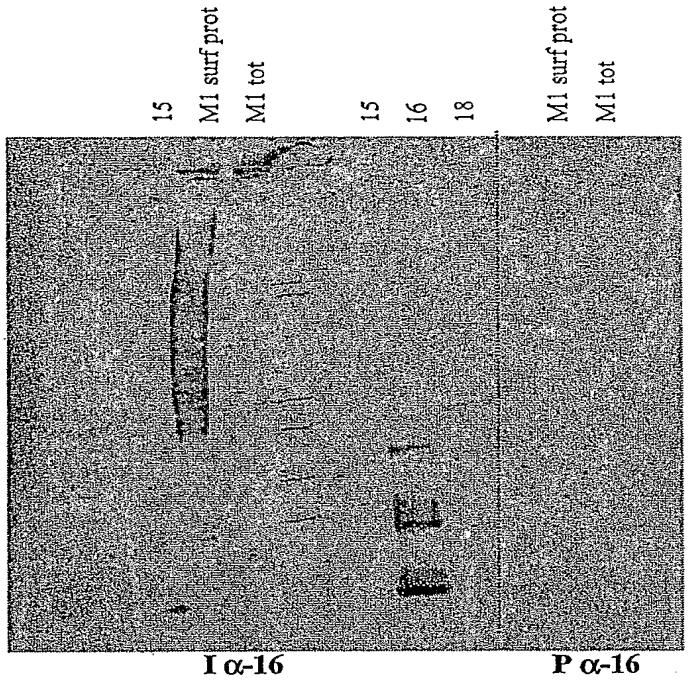
P α-15

LEGENDA:

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 91

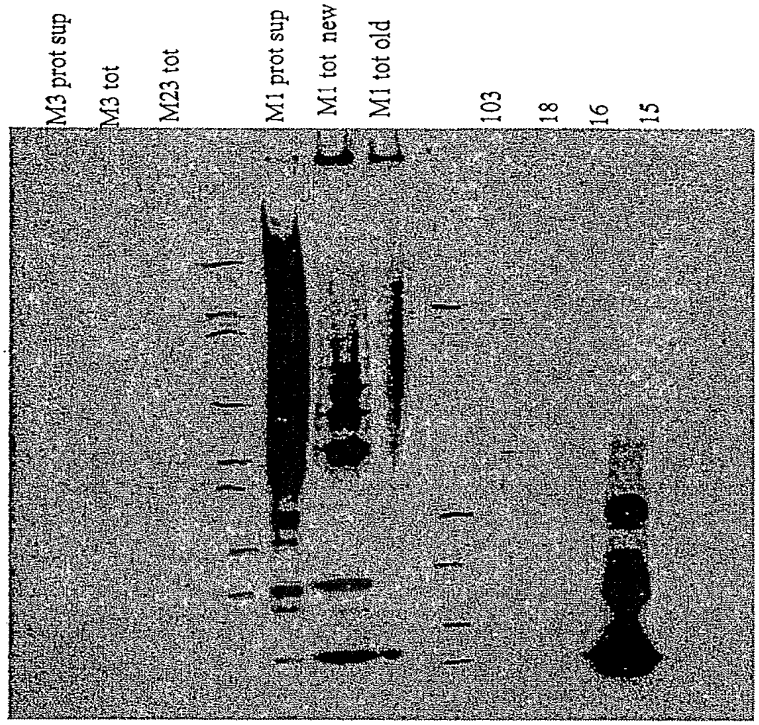
PCT/US05/27239 242/487



LEGEND:

M1 tot: total extract (M1)
M1 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

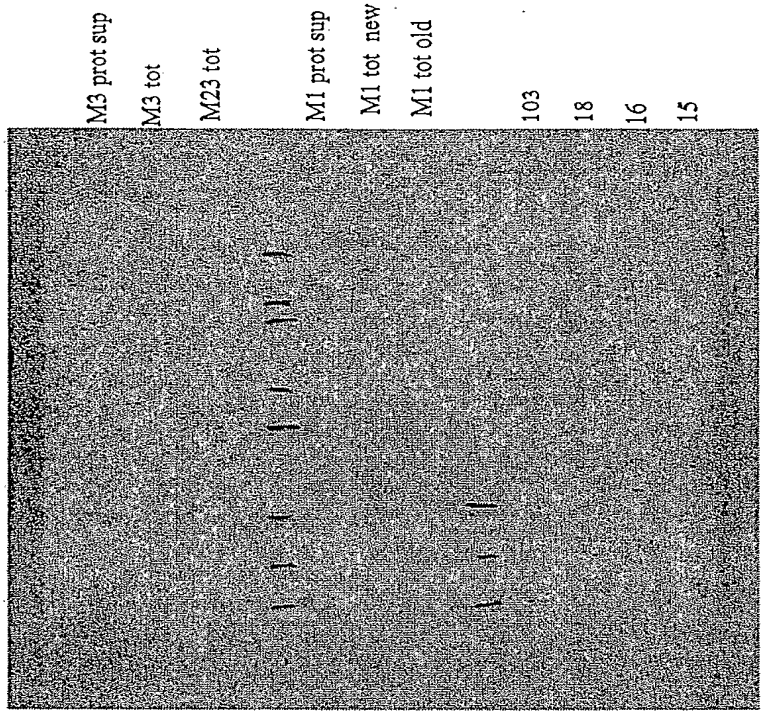
Figure 92



LEGENDA:

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 93

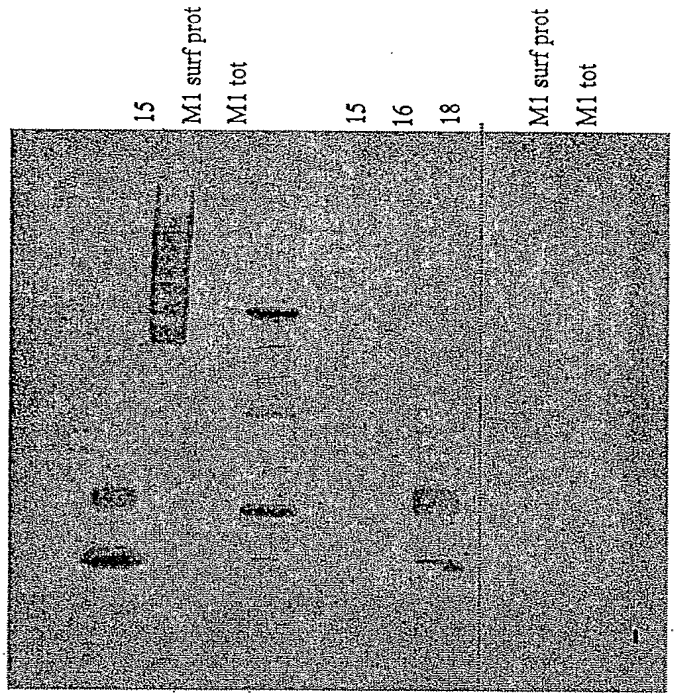


P α-16

LEGENDA:

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

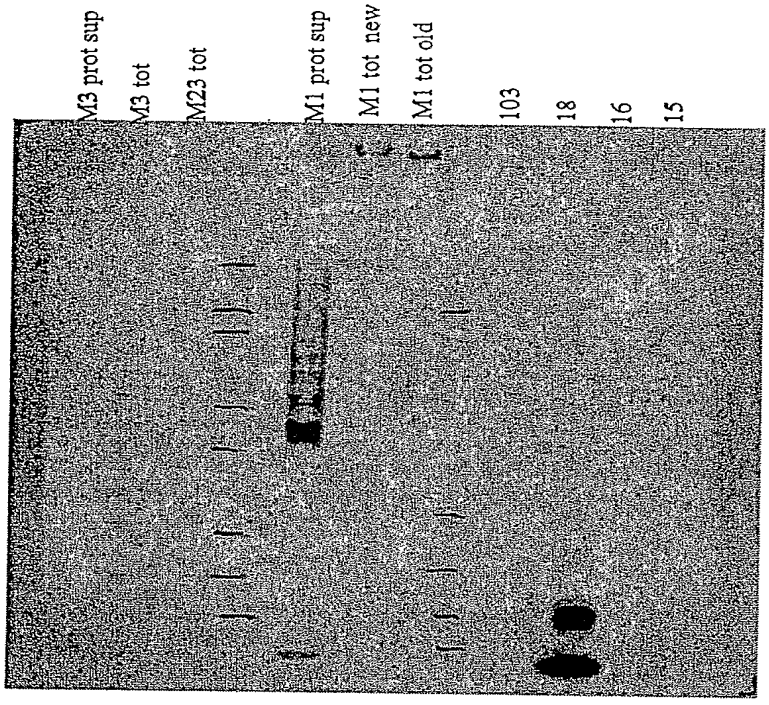
Figure 94



LEGEND:

M1 tot: total extract (M1)
M1 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 95

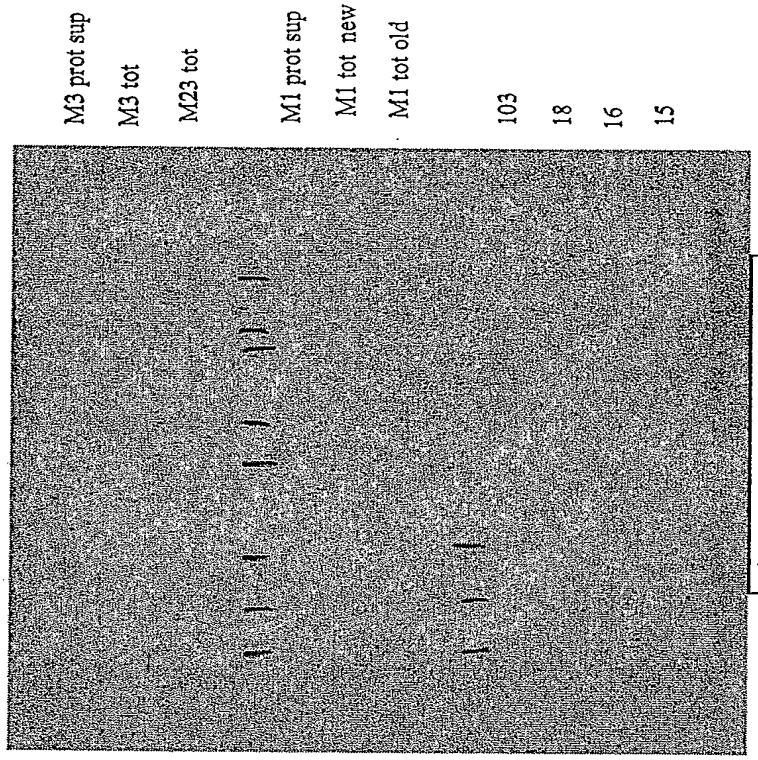


I α-18

LEGENDA:

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 96



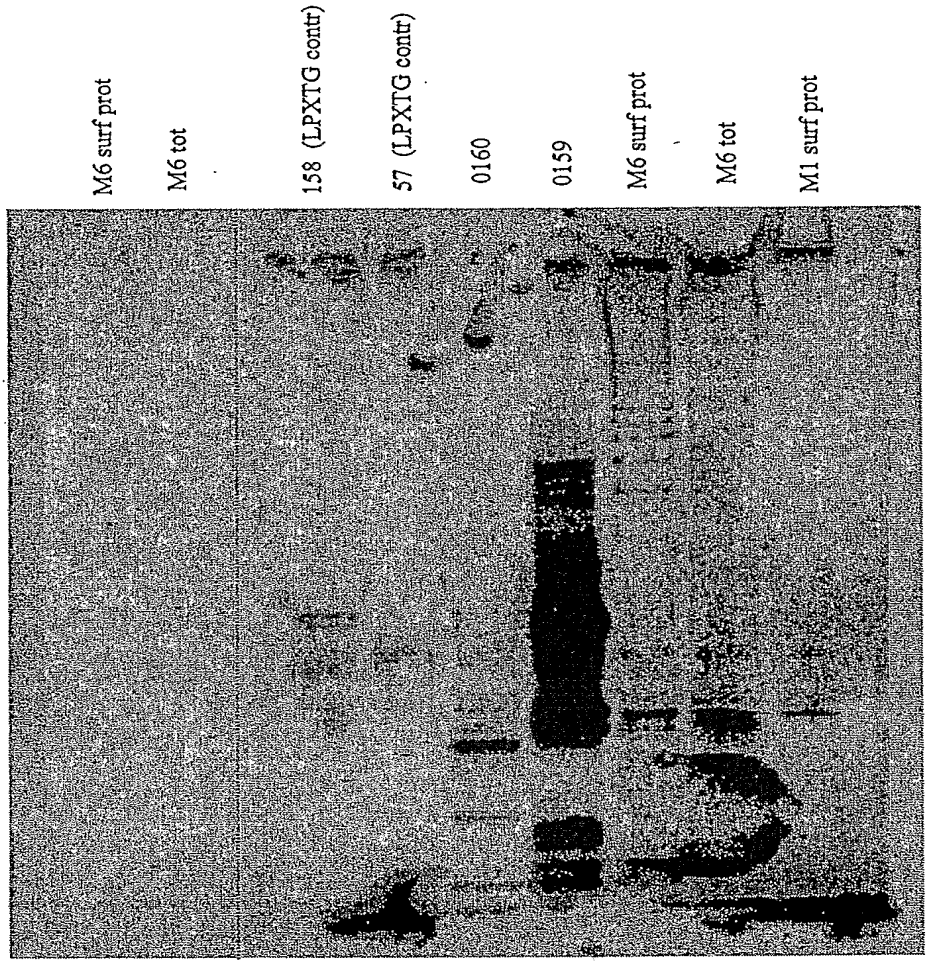
LEGENDA:

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

P α-18

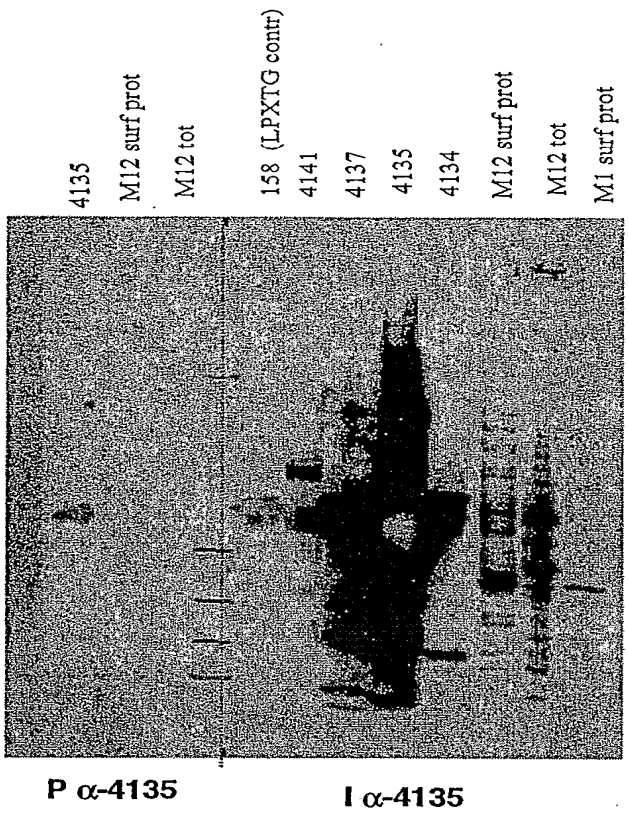
Figure 97

Figure 98



LEGEND:

M6 tot: total extract (M6)
M6 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

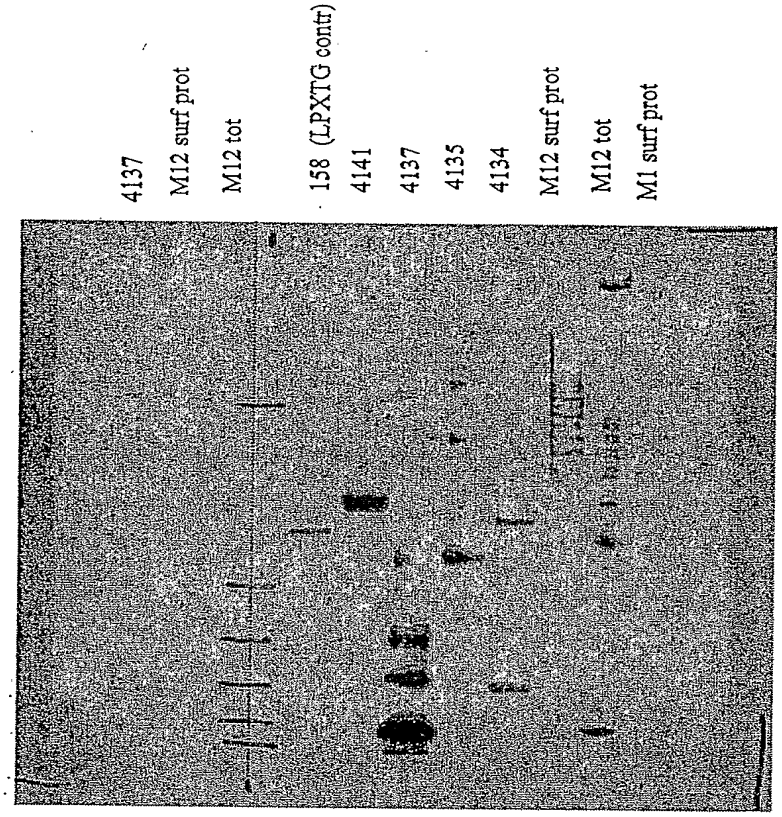


LEGEND:

M12 tot: total extract (M12)
M12 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 99

250/487



LEGEND:

M12 tot: total extract (M12)
M12 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

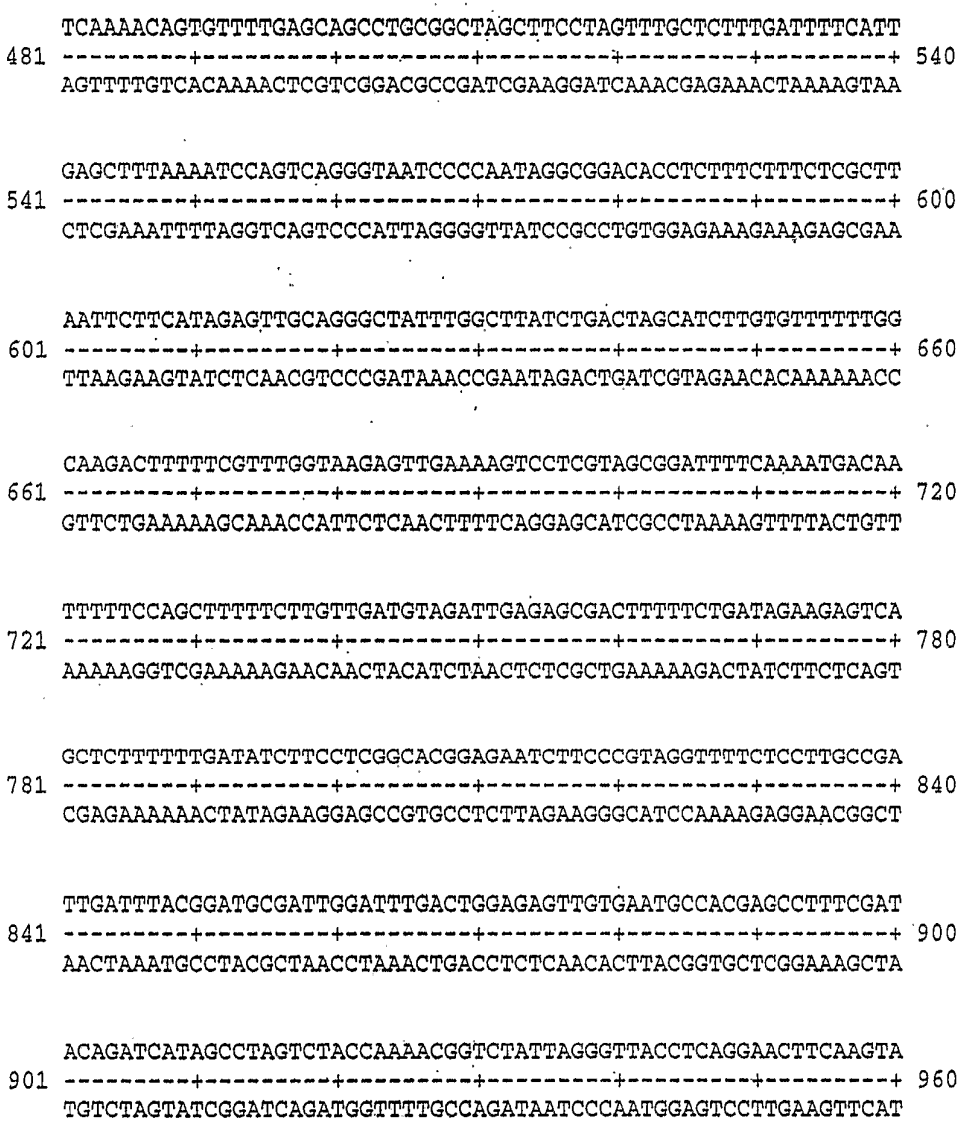
P α-4137 **I α-4137**

Figure 100

FIGURE 101


GACAAGCTTCCTTATACGACCGCTTTCTATATCGGACTCTTCCAAGTTCTTGCTCTTTTA
1 60
CTGTTCGAAGGAATATGCTGGCGAAAGATATAGCCTGAGAAGGTTCAAGAACGAGAAAAT
CCAGGGACTAGCCGTTCAGGTGCAACGATTGTCGGTGGTTTGTTAAATGGAACCAGTCGT
61 120
GGTCCCTGATCGGCAAGTCCACGTTGCTAACAGCCACCAAACAATTTACCTTGGTCAGCA
TCAGTTGTGACAGAATTTACCTTCTATCTTGGGATTCCCGTTATGTTTGGAGCTAGTGCC
121 180
AGTCAACACTGTCTTAAATGGAAGATAGAACCCTAAGGGCAATACAAACCTCGATCACGG
TTAAAGATTTTCAAATTTGTGAAAGCCGGAGAACTCTTGAGCTTTGGGCAATTGTTTTTG
181 240
AATTTCTAAAAGTTTAAACACTTTCGGCCTCTTGAGAACTCGAAACCCGTTAACAAAAAC
CTCTTGGTCGCGATGGGAGTAGCTTTTGCGGTCAGCATGGTGGCTATTCGCTTCTTGACC
241 300
GAGAACCAGCGCTACCCTCATCGAAAACGCCAGTCGTACCACCGATAAGCGAAGAACTGG
AGCTATGTGAAAAAACACGACTTCACCCTTTTTGGTAAATACCGTATCGTGCTTGGTAGT
301 360
TCGATACACTTTTTTGTGCTGAAGTGGGAAAAACCATTTATGGCATAGCACGAACCATCA
GTTTTGCTACTTTACAGTTTTGTCCGTTTATTTGTATAAGAAAAACCTTGAAGGGGTAAC
361 420
CAAAACGATGAAATGTCAAAACAGGCAAATAAACATATTCTTTTTGGAACTTCCCCATTG
TCTTCAAGGTTTTATACTCTTAGAAAATCTCTTCAAACCGCGTCAGCTTTATCTGCAACC
421 480
AGAAGTTCCAAAATATGAGAATCTTTTAGAGAAGTTTGGCGCAGTCGAAATAGACGTTGG

Figure 101A



WO 2006/078318 PCT/US2005/027239

Figure 101B


AATCAGCACCAGTAAAAACGCCCATTTGATGAAGACGTTCTACTGTCTTTTTTCCTACTC
961 ---------+---------+---------+---------+---------+---------+ 1020
TTAGTCGTGGTCATTTTTGCGGGTAAACTACTTCTGCAAGATGACAGAAAAAAGGATGAG
CATGAAATTTGGAAATATCCATTTGTTTGAGAAAATCCTCAGCCTGTTCAGGTAGAATCA
1021 ---------+---------+---------+---------+---------+---------+ 1080
GTACTTTAAACCTTTATAGGTAAACAAACTCTTTTAGGAGTCGGACAAGTCCATCTTAGT
CTGTCAAACCATGTGGTTTTTGATAATCACTCGCCATTTTAGCTAAGAATTTGTTGTAAG
1081 ---------+---------+---------+---------+---------+---------+ 1140
GACAGTTTGGTACACCAAAAACTATTAGTGAGCGGTAAAATCGATTCTTAAACAACATTC
AAACGCCTGCGGAAGCAGTTAGATGGAGTTCTTTCCAGATATCTTTTTGAATGAGGCGAG
1141 ---------+---------+---------+---------+---------+---------+ 1200
TTTGCGGACGCCTTCGTCAATCTACCTCAAGAAAGGTCTATAGAAAAACTTACTCCGCTC
CAATTTTGACCGCTGACTTGATACCGAGTTTATTTTCTGTCACATCCAAATAGGCTTCGT
1201 ---------+---------+---------+---------+---------+---------+ 1260
GTTAAAACTGGCGACTGAACTATGGCTCAAATAAAAGACAGTGTAGGTTTATCCGAAGCA
CAATGCTCATGGGTTCAATCAAATCTGTATAGCGCTTAAAAATAGCTCGAATCCGGAGTC
1261 ---------+---------+---------+---------+---------+---------+ 1320
GTTACGAGTACCCAAGTTAGTTTAGACATATCGCGAATTTTTATCGAGCTTAGGCCTCAG
CCACAGACTTGTATTTCTCATAATTCCCTGAGATAAAGACAGCCTGGGGACAACGTTCAT
1321 ---------+---------+---------+---------+---------+---------+ 1380
GGTGTCTGAACATAAAGAGTATTAAGGGACTCTATTTCTGTCGGACCCCTGTTGCAAGTA
AAGCTTCCTTGGAACTCATGGCAGAATGGACACCAAAAGCTCTTGCCTCATAACTACAGG
1381 ---------+---------+---------+---------+---------+---------+ 1440
TTCGAAGGAACCTTGAGTACCGTCTTACCTGTGGTTTTCGAGAACGGAGTATTGATGTCC

Figure 101C

```
    TAGAAACGACTCCCCGTCCACCTGTTTGCCGAGGGTCGCTTCCAATAATGACAGGTTTTC
1441 ---------+---------+---------+---------+---------+---------+ 1500
    ATCTTTGCTGAGGGGCAGGTGGACAAACGGCTCCCAGCGAAGGTTATTACTGTCCAAAAG

    CTCTGAGTTTAGGATTATCCCTGATTTCCACTGCAGCAAAAAAGGCATCCATGTCAATAT
1501 ---------+---------+---------+---------+---------+---------+ 1560
    GAGACTCAAATCCTAATAGGGACTAAAGGTGACGTCGTTTTTTCCGTAGGTACAGTTATA

    GGATGATTTTTCTTGACAAATCATTTAACAAAGGAAAAATCAACATGCCTAGCACCTTTT
1561 ---------+---------+---------+---------+---------+---------+ 1620
    CCTACTAAAAAGAACTGTTTAGTAAATTGTTTCCTTTTTAGTTGTACGGATCGTGGAAAA

    TATACTCTTCGAAAATCTCTTCAAACCACGTCAGCTTCCATCTGCAACCTCAAAACAGTA
1621 ---------+---------+---------+---------+---------+---------+ 1680
    ATATGAGAAGCTTTTAGAGAAGTTTGGTGCAGTCGAAGGTAGACGTTGGAGTTTTGTCAT

    TTTTGAGCTGACTTCGTCAGTTCTATTTACAACCTCAAAGCAGTGCTTTGAGCAGCCTGC
1681 ---------+---------+---------+---------+---------+---------+ 1740
    AAAACTCGACTGAAGCAGTCAAGATAAATGTTGGAGTTTCGTCACGAAACTCGTCGGACG

    GGCTAGTTTCCTAGTTTGCTTTTCGATTTCCATTGAGTGTAACTGCTTATTTTCTTTTAT
1741 ---------+---------+---------+---------+---------+---------+ 1800
    CCGATCAAAGGATCAAACGAAAAGCTAAAGGTAACTCACATTGACGAATAAAAGAAAATA

    TATACCCTTTTTTCTGAAAAAAAGAAAAAAGGACTTTATTTTTTTCAAAAATATAATACA
1801 ---------+---------+---------+---------+---------+---------+ 1860
    ATATGGGAAAAAAGACTTTTTTTCTTTTTTCCTGAAATAAAAAAAGTTTTTATATTATGT

    GTTTGAAATAAAATATAGACTGTTTTAGAAAAGAAAGTGTAAAAATAGGAATTTTTCACT
1861 ---------+---------+---------+---------+---------+---------+ 1920
    CAAACTTTATTTTATATCTGACAAAATCTTTTCTTTCACATTTTTATCCTTAAAAAGTGA
```

Figure 101D

```
      TGTTGAAATCGGTTACTTTATGGTATACTTGTCTCATGAATGTAACAGATGACTGTTACT
1921  ---------+---------+---------+---------+---------+---------+ 1980
      ACAACTTTAGCCAATGAAATACCATATGAACAGAGTACTTACATTGTCTACTGACAATGA

      AGAAAAAAGAGGACATTAATATGGTTGTTAAGACAGTTGTTGAAGCACAAGATATTTTTG
1981  ---------+---------+---------+---------+---------+---------+ 2040
      TCTTTTTTCTCCTGTAATTATACCAACAATTCTGTCAACAACTTCGTGTTCTATAAAAAC

c                     M  V  V  K  T  V  V  E  A  Q  D  I  F  D -  this orf is the homologue of sp0459, a formate
                                                                   acetyltransferase (pfl). It is out of the
                                                                   pilus locus
      ACAAAGCTTGGGAAGGCTTCAAAGGCGTAGATTGGAAAGAAAAAGCAAGTGTATCACGCT
2041  ---------+---------+---------+---------+---------+---------+ 2100
      TGTTTCGAACCCTTCCGAAGTTTCCGCATCTAACCTTTCTTTTTCGTTCACATAGTGCGA

c       K  A  W  E  G  F  K  G  V  D  W  K  E  K  A  S  V  S  R  F -

      TTGTACAAGCTAACTACACACCTTATGATGGAGACGAAAGCTTCCTTGCAGGACCAACAG
2101  ---------+---------+---------+---------+---------+---------+ 2160
      AACATGTTCGATTGATGTGTGGAATACTACCTCTGCTTTCGAAGGAACGTCCTGGTTGTC

c       V  Q  A  N  Y  T  P  Y  D  G  D  E  S  F  L  A  G  P  T  E -

      AGCGTTCACTTCACATCAAGAAAATTGTAGAAGAAACTAAAGCACACTACGAAGAAACTC
2161  ---------+---------+---------+---------+---------+---------+ 2220
      TCGCAAGTGAAGTGTAGTTCTTTTAACATCTTCTTTGATTTCGTGTGATGCTTCTTTGAG

c       R  S  L  H  I  K  K  I  V  E  E  T  K  A  H  Y  E  E  T  R -

      GTTTCCCAATGGACACTCGTCCAACATCTATCGCTGATATCCCTGCTGGATTTATCGACA
2221  ---------+---------+---------+---------+---------+---------+ 2280
      CAAAGGGTTACCTGTGAGCAGGTTGTAGATAGCGACTATAGGGACGACCTAAATAGCTGT

c       F  P  M  D  T  R  P  T  S  I  A  D  I  P  A  G  F  I  D  K -

      AAGAAAATGAAGTTATCTTTGGTATCCAAAATGATGAACTCTTCAAATTGAACTTCATGC
2281  ---------+---------+---------+---------+---------+---------+ 2340
      TTCTTTTACTTCAATAGAAACCATAGGTTTTACTACTTGAGAAGTTTAACTTGAAGTACG
```

Figure 101E

```
c         E  N  E  V  I  F  G  I  Q  N  D  E  L  F  K  L  N  F  M  P -

      CAAAAGGTGGTATCCGTATGGCTGAAACTACTTTAAAAGAAAATGGATACGAACCAGACC
2341  ---------+---------+---------+---------+---------+---------+ 2400
      GTTTTCCACCATAGGCATACCGACTTTGATGAAATTTTCTTTTACCTATGCTTGGTCTGG

c         K  G  G  I  R  M  A  E  T  T  L  K  E  N  G  Y  E  P  D  P -

      CAGCTGTTCACGAAATCTTCACTAAATATGTAACAACAGTTAACGACGGTATTTTCCGTG
2401  ---------+---------+---------+---------+---------+---------+ 2460
      GTCGACAAGTGCTTTAGAAGTGATTTATACATTGTTGTCAATTGCTGCCATAAAAGGCAC

c         A  V  H  E  I  F  T  K  Y  V  T  T  V  N  D  G  I  F  R  A -

      CCTACACTTCAAATATTCGTCGCGCTCGTCATGCACACACTGTAACTGGTCTTCCAGATG
2461  ---------+---------+---------+---------+---------+---------+ 2520
      GGATGTGAAGTTTATAAGCAGCGCGAGCAGTACGTGTGTGACATTGACCAGAAGGTCTAC

c         Y  T  S  N  I  R  R  A  R  H  A  H  T  V  T  G  L  P  D  A -

      CATACTCACGCGGACGTATCATCGGTGTTTACGCACGTCTTGCTCTTTACGGTGCAGACT
2521  ---------+---------+---------+---------+---------+---------+ 2580
      GTATGAGTGCGCCTGCATAGTAGCCACAAATGCGTGCAGAACGAGAAATGCCACGTCTGA

c         Y  S  R  G  R  I  I  G  V  Y  A  R  L  A  L  Y  G  A  D  Y -

      ACTTGATGCAAGAAAAAGTAAACGACTGGAATGCAATCAAAGAAATCGATGAAGAAACAA
2581  ---------+---------+---------+---------+---------+---------+ 2640
      TGAACTACGTTCTTTTTCATTTGCTGACCTTACGTTAGTTTCTTTAGCTACTTCTTTGTT

c         L  M  Q  E  K  V  N  D  W  N  A  I  K  E  I  D  E  E  T  I -

      TCCGTCTTCGTGAAGAAGTAAACCTTCAATACCAAGCATTGCAACAAGTTGTTCGCCTGG
2641  ---------+---------+---------+---------+---------+---------+ 2700
      AGGCAGAAGCACTTCTTCATTTGGAAGTTATGGTTCGTAACGTTGTTCAACAAGCGGACC

c         R  L  R  E  E  V  N  L  Q  Y  Q  A  L  Q  Q  V  V  R  L  G -

      GTGACCTTTACGGGGTTGATGTTCGCAAACCAGCGATGAACGTGAAAGAAGCAATCCAAT
2701  ---------+---------+---------+---------+---------+---------+ 2760
```

Figure 101F

```
          CACTGGAAATGCCCCAACTACAAGCGTTTGGTCGCTACTTGCACTTTCTTCGTTAGGTTA

c           D  L  Y  G  V  D  V  R  K  P  A  M  N  V  K  E  A  I  Q  W -

          GGGTTAACATTGCTTTCATGGCTGTCTGCCGTGTGATTAACGGTGCTGCTACATCTCTAG
     2761 ---------+---------+---------+---------+---------+---------+ 2820
          CCCAATTGTAACGAAAGTACCGACAGACGGCACACTAATTGCCACGACGATGTAGAGATC

c           V  N  I  A  F  M  A  V  C  R  V  I  N  G  A  A  T  S  L  G -

          GTCGTGTACCAATCGTATTGGACATCTTTGCAGAACGTGACCTTGCTCGTGGTACATTTA
     2821 ---------+---------+---------+---------+---------+---------+ 2880
          CAGCACATGGTTAGCATAACCTGTAGAAACGTCTTGCACTGGAACGAGCACCATGTAAAT

c           R  V  P  I  V  L  D  I  F  A  E  R  D  L  A  R  G  T  F  T -

          CTGAATCAGAAATCCAAGAATTCGTTGATGATTTCGTTATGAAACTTCGTACAGTTAAAT
     2881 ---------+---------+---------+---------+---------+---------+ 2940
          GACTTAGTCTTTAGGTTCTTAAGCAACTACTAAAGCAATACTTTGAAGCATGTCAATTTA

c           E  S  E  I  Q  E  F  V  D  D  F  V  M  K  L  R  T  V  K  F -

          TTGCTCGTACCAAAGCTTATGACCAATTGTACTCAGGTGACCCAACCTTTATCACAACTT
     2941 ---------+---------+---------+---------+---------+---------+ 3000
          AACGAGCATGGTTTCGAATACTGGTTAACATGAGTCCACTGGGTTGGAAATAGTGTTGAA

c           A  R  T  K  A  Y  D  Q  L  Y  S  G  D  P  T  F  I  T  T  S -

          CTATGGCTGGTATGGGTAACGACGGTCGTCACCGTGTTACTAAGATGGACTACCGTTTCT
     3001 ---------+---------+---------+---------+---------+---------+ 3060
          GATACCGACCATACCCATTGCTGCCAGCAGTGGCACAATGATTCTACCTGATGGCAAAGA

c           M  A  G  M  G  N  D  G  R  H  R  V  T  K  M  D  Y  R  F  L -

          TGAACACTCTTGACAACATCGGTAACTCACCAGAACCAAACTTGACAGTTCTTTGGACTG
     3061 ---------+---------+---------+---------+---------+---------+ 3120
          ACTTGTGAGAACTGTTGTAGCCATTGAGTGGTCTTGGTTTGAACTGTCAAGAAACCTGAC

c           N  T  L  D  N  I  G  N  S  P  E  P  N  L  T  V  L  W  T  D -
```

Figure 101G

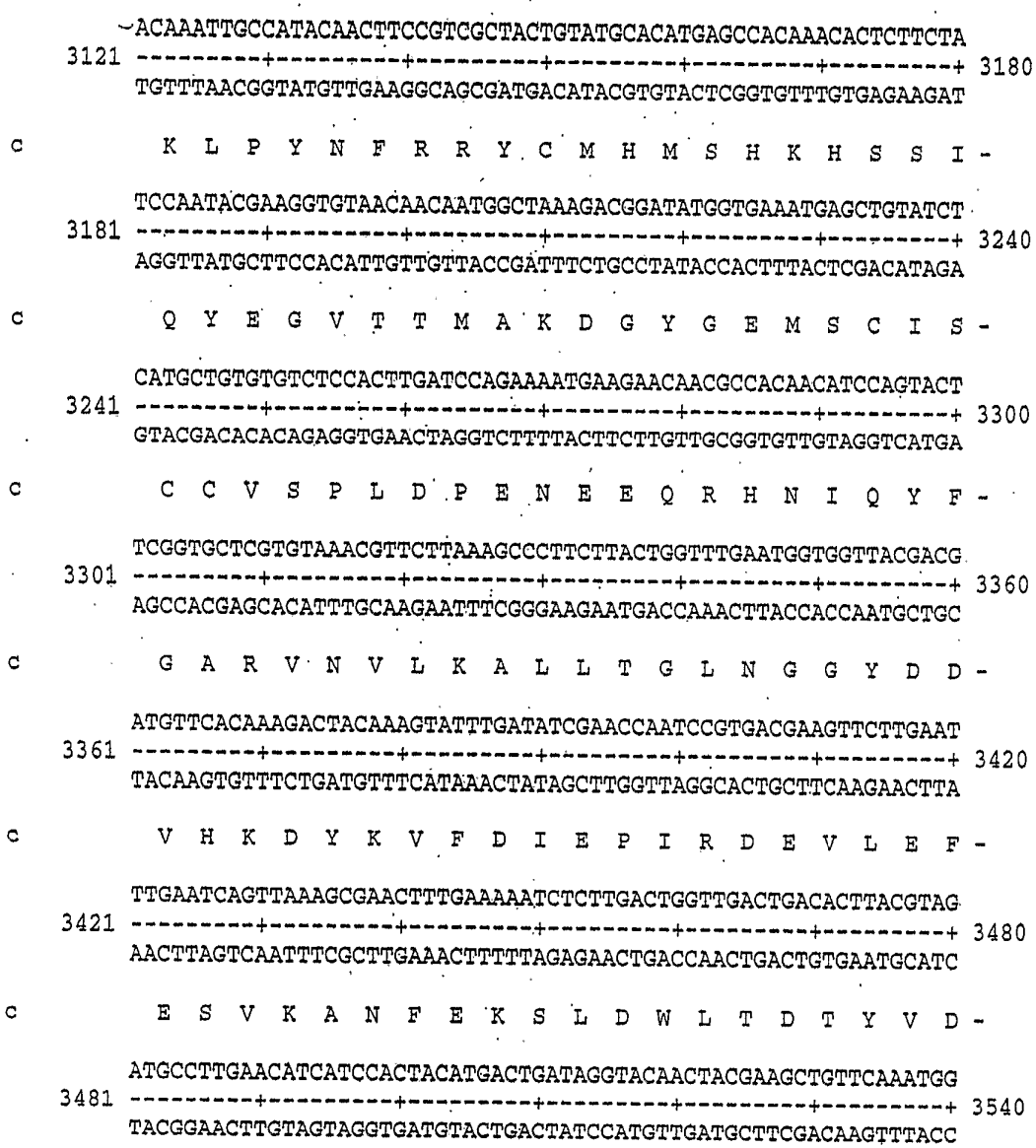

ACAAATTGCCATACAACTTCCGTCGCTACTGTATGCACATGAGCCACAAACACTCTTCTA
3121 + 3180
TGTTTAACGGTATGTTGAAGGCAGCGATGACATACGTGTACTCGGTGTTTGTGAGAAGAT
c K L P Y N F R R Y C M H M S H K H S S I -
TCCAATACGAAGGTGTAACAACAATGGCTAAAGACGGATATGGTGAAATGAGCTGTATCT
3181 + 3240
AGGTTATGCTTCCACATTGTTGTTACCGATTTCTGCCTATACCACTTTACTCGACATAGA
c Q Y E G V T T M A K D G Y G E M S C I S -
CATGCTGTGTGTCTCCACTTGATCCAGAAAATGAAGAACAACGCCACAACATCCAGTACT
3241 + 3300
GTACGACACACAGAGGTGAACTAGGTCTTTTACTTCTTGTTGCGGTGTTGTAGGTCATGA
c C C V S P L D P E N E E Q R H N I Q Y F -
TCGGTGCTCGTGTAAACGTTCTTAAAGCCCTTCTTACTGGTTTGAATGGTGGTTACGACG
3301 + 3360
AGCCACGAGCACATTTGCAAGAATTTCGGGAAGAATGACCAAACTTACCACCAATGCTGC
c G A R V N V L K A L L T G L N G G Y D D -
ATGTTCACAAAGACTACAAAGTATTTGATATCGAACCAATCCGTGACGAAGTTCTTGAAT
3361 + 3420
TACAAGTGTTTCTGATGTTTCATAAACTATAGCTTGGTTAGGCACTGCTTCAAGAACTTA
c V H K D Y K V F D I E P I R D E V L E F -
TTGAATCAGTTAAAGCGAACTTTGAAAAATCTCTTGACTGGTTGACTGACACTTACGTAG
3421 + 3480
AACTTAGTCAATTTCGCTTGAAACTTTTTAGAGAACTGACCAACTGACTGTGAATGCATC
c E S V K A N F E K S L D W L T D T Y V D -
ATGCCTTGAACATCATCCACTACATGACTGATAGGTACAACTACGAAGCTGTTCAAATGG
3481 + 3540
TACGGAACTTGTAGTAGGTGATGTACTGACTATCCATGTTGATGCTTCGACAAGTTTACC

Figure 101H

```
c          A  L  N  I  I  H  Y  M  T  D  R  Y  N  Y  E  A  V  Q  M  A -

       CCTTCTTGCCAACTAAACAACGTGCCAACATGGGATTCGGTATCTGTGGATTTGCTAACA
3541   ---------+---------+---------+---------+---------+---------+ 3600
       GGAAGAACGGTTGATTTGTTGCACGGTTGTACCCTAAGCCATAGACACCTAAACGATTGT

c          F  L  P  T  K  Q  R  A  N  M  G  F  G  I  C  G  F  A  N  T -

       CTGTTGATACATTGTCAGCTATCAAATACGCTACAGTTAAACCAATCCGTGACGAAGATG
3601   ---------+---------+---------+---------+---------+---------+ 3660
       GACAACTATGTAACAGTCGATAGTTTATGCGATGTCAATTTGGTTAGGCACTGCTTCTAC

c          V  D  T  L  S  A  I  K  Y  A  T  V  K  P  I  R  D  E  D  G -

       GCTACATCTACGATTACGAAACAATCGGTGACTACCCACGCTGGGGTGAAGATGACCCAC
3661   ---------+---------+---------+---------+---------+---------+ 3720
       CGATGTAGATGCTAATGCTTTGTTAGCCACTGATGGGTGCGACCCCACTTCTACTGGGTG

c          Y  I  Y  D  Y  E  T  I  G  D  Y  P  R  W  G  E  D  D  P  R -

       GTTCAAACGAATTGGCAGAATGGTTGATCGAAGCTTACACAACTCGTCTACGTAGCCACA
3721   ---------+---------+---------+---------+---------+---------+ 3780
       CAAGTTTGCTTAACCGTCTTACCAACTAGCTTCGAATGTGTTGAGCAGATGCATCGGTGT

c          S  N  E  L  A  E  W  L  I  E  A  Y  T  T  R  L  R  S  H  K -

       AACTATACAAAGACGCAGAAGCTACAGTATCACTTTTGACAATCACATCTAACGTTGCTT
3781   ---------+---------+---------+---------+---------+---------+ 3840
       TTGATATGTTTCTGCGTCTTCGATGTCATAGTGAAAACTGTTAGTGTAGATTGCAACGAA

c          L  Y  K  D  A  E  A  T  V  S  L  L  T  I  T  S  N  V  A  Y -

       ACTCTAAACAAACTGGTAACTCACCAGTTCACAAAGGTGTATACCTCAACGAAGATGGTT
3841   ---------+---------+---------+---------+---------+---------+ 3900
       TGAGATTTGTTTGACCATTGAGTGGTCAAGTGTTTCCACATATGGAGTTGCTTCTACCAA

c          S  K  Q  T  G  N  S  P  V  H  K  G  V  Y  L  N  E  D  G  S -

       CTGTGAACTTGTCTAAACTTGAATTCTTCTCACCAGGTGCTAACCCATCTAACAAAGCTA
3901   ---------+---------+---------+---------+---------+---------+ 3960
```

Figure 101I

```
          GACACTTGAACAGATTTGAACTTAAGAAGAGTGGTCCACGATTGGGTAGATTGTTTCGAT

c           V  N  L  S  K  L  E  F  F  S  P  G  A  N  P  S  N  K  A  K -

          AAGGTGGTTGGTTGCAAAACTTGAACTCACTTTCTAGCCTTGACTTTAGTTATGCAGCTG
     3961 ---------+---------+---------+---------+---------+---------+ 4020
          TTCCACCAACCAACGTTTTGAACTTGAGTGAAAGATCGGAACTGAAATCAATACGTCGAC

c           G  G  W  L  Q  N  L  N  S  L  S  S  L  D  F  S  Y  A  A  D -

          ACGGTATCTCATTGACTACACAAGTATCACCTCGCGCTCTTGGTAAGACTCGTGATGAAC
     4021 ---------+---------+---------+---------+---------+---------+ 4080
          TGCCATAGAGTAACTGATGTGTTCATAGTGGAGCGCGAGAACCATTCTGAGCACTACTTG

c           G  I  S  L  T  T  Q  V  S  P  R  A  L  G  K  T  R  D  E  Q -

          AAGTTGATAACTTGGTAACAATCCTTGATGGTTACTTCGAAAACGGTGGACAACACGTTA
     4081 ---------+---------+---------+---------+---------+---------+ 4140
          TTCAACTATTGAACCATTGTTAGGAACTACCAATGAAGCTTTTGCCACCTGTTGTGCAAT

c           V  D  N  L  V  T  I  L  D  G  Y  F  E  N  G  G  Q  H  V  N -

          ACTTGAACGTTATGGACTTGAACGATGTTTACGAAAAAATCATGTCAGGCGAAGACGTTA
     4141 ---------+---------+---------+---------+---------+---------+ 4200
          TGAACTTGCAATACCTGAACTTGCTACAAATGCTTTTTTAGTACAGTCCGCTTCTGCAAT

c           L  N  V  M  D  L  N  D  V  Y  E  K  I  M  S  G  E  D  V  I -

          TCGTACGTATCTCTGGATACTGTGTAAACACTAAATACCTCACTCCAGAACAAAAAACTG
     4201 ---------+---------+---------+---------+---------+---------+ 4260
          AGCATGCATAGAGACCTATGACACATTTGTGATTTATGGAGTGAGGTCTTGTTTTTTGAC

c           V  R  I  S  G  Y  C  V  N  T  K  Y  L  T  P  E  Q  K  T  E -

          AATTGACACAACGTGTCTTCCACGAAGTTCTTTCAATGGATGACGCCTTGGATGCATTGA
     4261 ---------+---------+---------+---------+---------+---------+ 4320
          TTAACTGTGTTGCACAGAAGGTGCTTCAAGAAAGTTACCTACTGCGGAACCTACGTAACT

c           L  T  Q  R  V  F  H  E  V  L  S  M  D  D  A  L  D  A  L  S -
```

Figure 101J

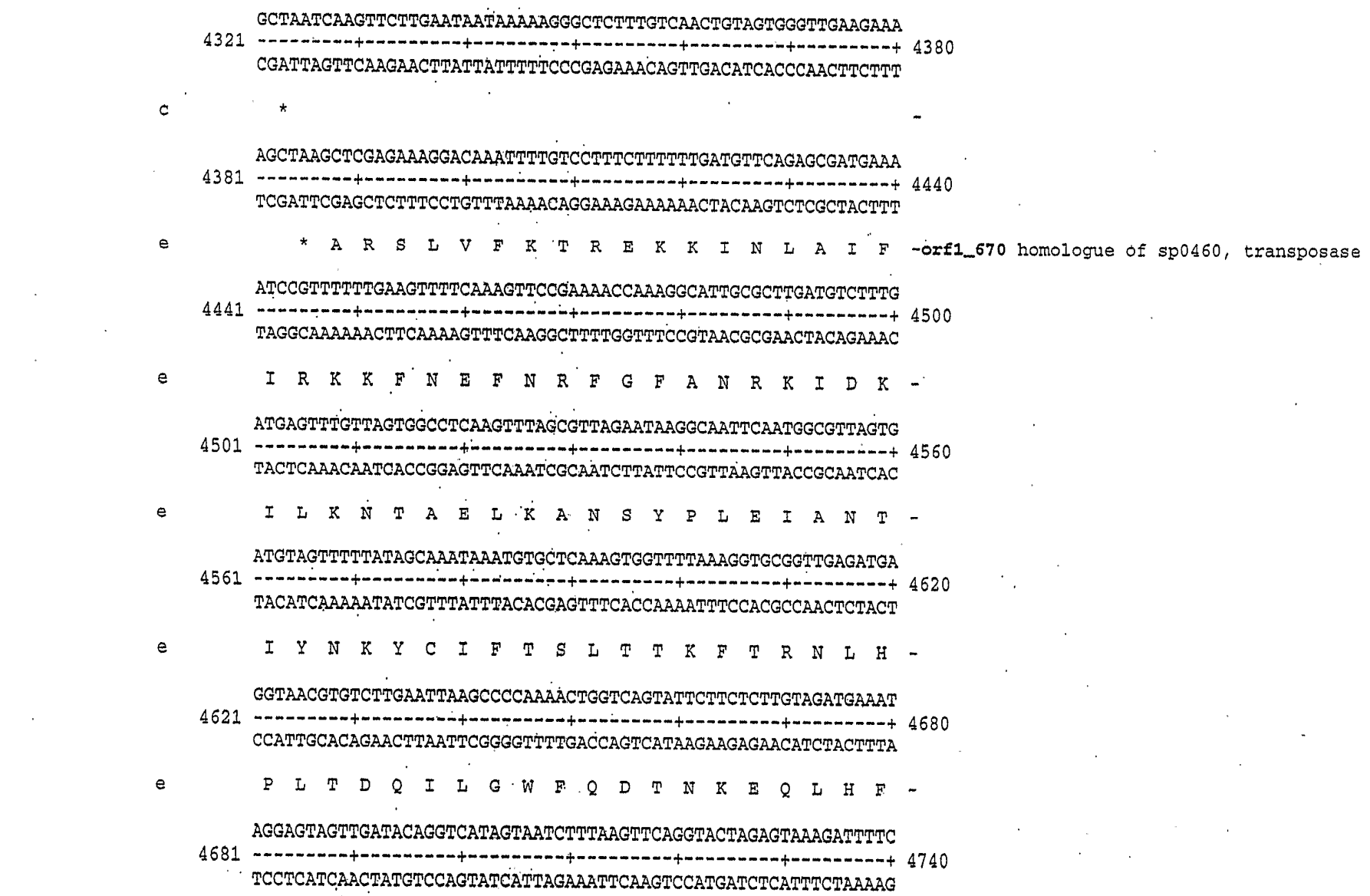

GCTAATCAAGTTCTTGAATAATAAAAAGGGCTCTTTGTCAACTGTAGTGGGTTGAAGAAA
4321 ---------+---------+---------+---------+---------+---------+ 4380
CGATTAGTTCAAGAACTTATTATTTTTCCCGAGAAACAGTTGACATCACCCAACTTCTTT
c *
-
AGCTAAGCTCGAGAAAGGACAAATTTTGTCCTTTCTTTTTTGATGTTCAGAGCGATGAAA
4381 ---------+---------+---------+---------+---------+---------+ 4440
TCGATTCGAGCTCTTTCCTGTTTAAAACAGGAAAGAAAAAACTACAAGTCTCGCTACTTT
e * A R S L V F K T R E K K I N L A I F -orf1_670 homologue of sp0460, transposase
ATCCGTTTTTTGAAGTTTTCAAAGTTCCGAAAACCAAAGGCATTGCGCTTGATGTCTTTG
4441 ---------+---------+---------+---------+---------+---------+ 4500
TAGGCAAAAAACTTCAAAAGTTTCAAGGCTTTTGGTTTCCGTAACGCGAACTACAGAAAC
e I R K K F N E F N R F G F A N R K I D K -
ATGAGTTTGTTAGTGGCCTCAAGTTTAGCGTTAGAATAAGGCAATTCAATGGCGTTAGTG
4501 ---------+---------+---------+---------+---------+---------+ 4560
TACTCAAACAATCACCGGAGTTCAAATCGCAATCTTATTCCGTTAAGTTACCGCAATCAC
e I L K N T A E L K A N S Y P L E I A N T -
ATGTAGTTTTTATAGCAAATAAATGTGCTCAAAGTGGTTTTAAAGGTGCGGTTGAGATGA
4561 ---------+---------+---------+---------+---------+---------+ 4620
TACATCAAAAATATCGTTTATTTACACGAGTTTCACCAAAATTTCCACGCCAACTCTACT
e I Y N K Y C I F T S L T T K F T R N L H -
GGTAACGTGTCTTGAATTAAGCCCCAAAACTGGTCAGTATTCTTCTCTTGTAGATGAAAT
4621 ---------+---------+---------+---------+---------+---------+ 4680
CCATTGCACAGAACTTAATTCGGGGTTTTGACCAGTCATAAGAAGAGAACATCTACTTTA
e P L T D Q I L G W F Q D T N K E Q L H F -
AGGAGTAGTTGATACAGGTCATAGTAATCTTTAAGTTCAGGTACTAGAGTAAAGATTTTC
4681 ---------+---------+---------+---------+---------+---------+ 4740
TCCTCATCAACTATGTCCAGTATCATTAGAAATTCAAGTCCATGATCTCATTTCTAAAAG

Figure 101K

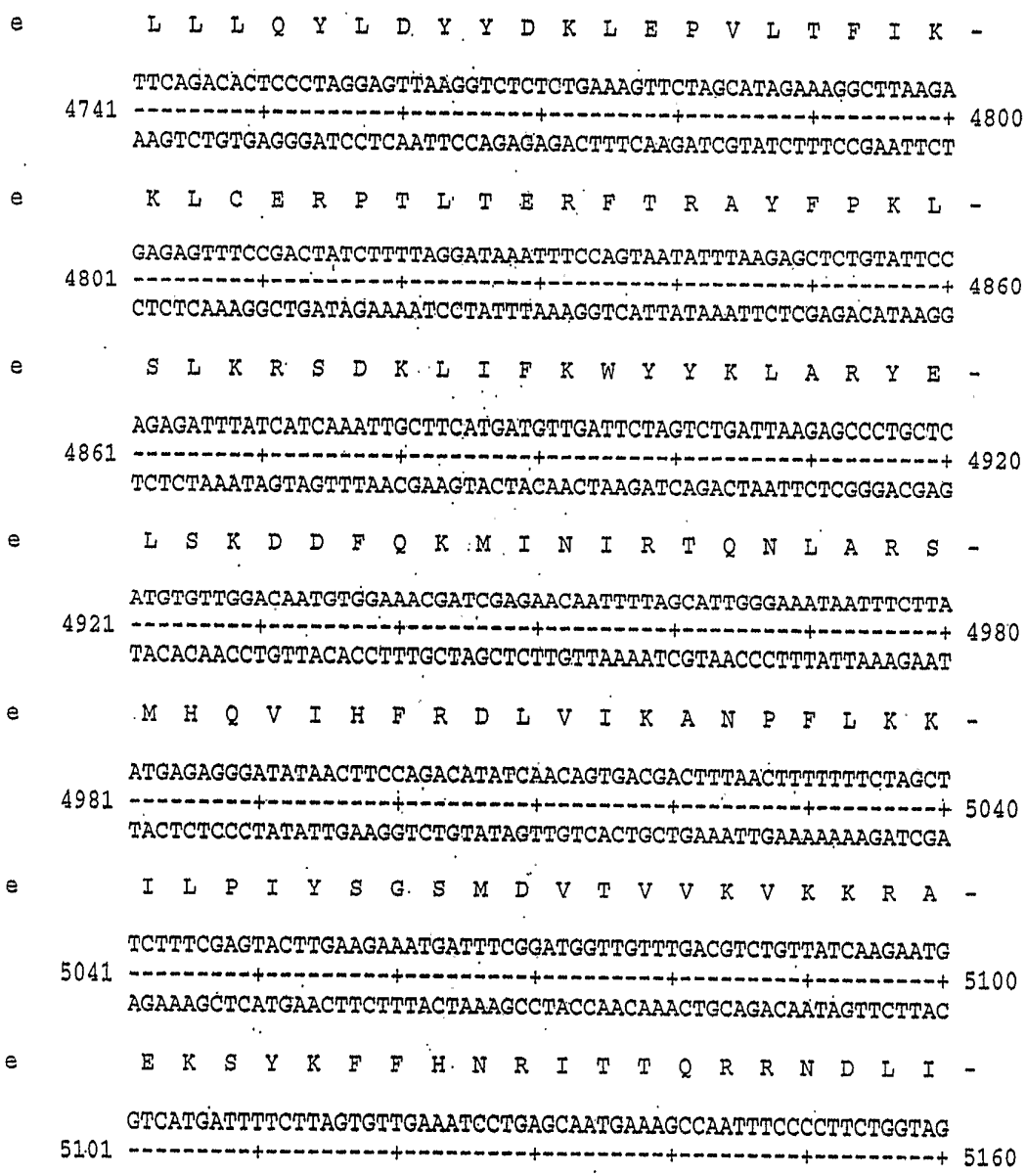

e L L L Q Y L D Y Y D K L E P V L T F I K -
TTCAGACACTCCCTAGGAGTTAAGGTCTCTCTGAAAGTTCTAGCATAGAAAGGCTTAAGA
4741 ---------+---------+---------+---------+---------+---------+ 4800
AAGTCTGTGAGGGATCCTCAATTCCAGAGAGACTTTCAAGATCGTATCTTTCCGAATTCT
e K L C E R P T L T E R F T R A Y F P K L -
GAGAGTTTCCGACTATCTTTTAGGATAAATTTCCAGTAATATTTAAGAGCTCTGTATTCC
4801 ---------+---------+---------+---------+---------+---------+ 4860
CTCTCAAAGGCTGATAGAAAATCCTATTTAAAGGTCATTATAAATTCTCGAGACATAAGG
e S L K R S D K L I F K W Y Y K L A R Y E -
AGAGATTTATCATCAAATTGCTTCATGATGTTGATTCTAGTCTGATTAAGAGCCCTGCTC
4861 ---------+---------+---------+---------+---------+---------+ 4920
TCTCTAAATAGTAGTTTAACGAAGTACTACAACTAAGATCAGACTAATTCTCGGGACGAG
e L S K D D F Q K M I N I R T Q N L A R S -
ATGTGTTGGACAATGTGGAAACGATCGAGAACAATTTTAGCATTGGGAAATAATTTCTTA
4921 ---------+---------+---------+---------+---------+---------+ 4980
TACACAACCTGTTACACCTTTGCTAGCTCTTGTTAAAATCGTAACCCTTTATTAAAGAAT
e M H Q V I H F R D L V I K A N P F L K K -
ATGAGAGGGATATAACTTCCAGACATATCAACAGTGACGACTTTAACTTTTTTTCTAGCT
4981 ---------+---------+---------+---------+---------+---------+ 5040
TACTCTCCCTATATTGAAGGTCTGTATAGTTGTCACTGCTGAAATTGAAAAAAAGATCGA
e I L P I Y S G S M D V T V V K V K K R A -
TCTTTCGAGTACTTGAAGAAATGATTTCGGATGGTTGTTTGACGTCTGTTATCAAGAATG
5041 ---------+---------+---------+---------+---------+---------+ 5100
AGAAAGCTCATGAACTTCTTTACTAAAGCCTACCAACAAACTGCAGACAATAGTTCTTAC
e E K S Y K F F H N R I T T Q R R N D L I -
GTCATGATTTTCTTAGTGTTGAAATCCTGAGCAATGAAAGCCAATTTCCCCTTCTGGTAG
5101 ---------+---------+---------+---------+---------+---------+ 5160

Figure 101L

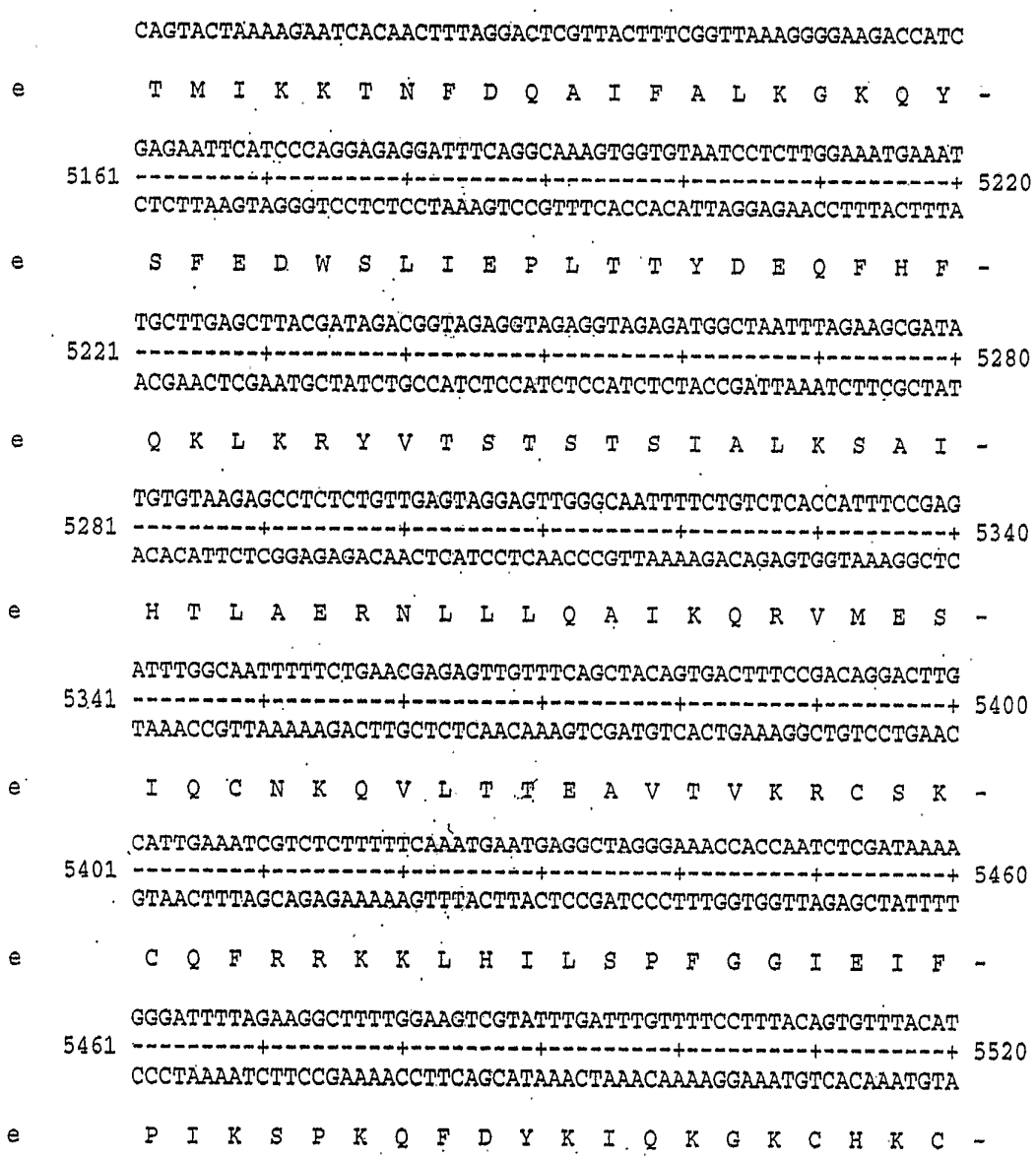

CAGTACTAAAAGAATCACAACTTTAGGACTCGTTACTTTCGGTTAAAGGGGAAGACCATC
e T M I K K T N F D Q A I F A L K G K Q Y -
GAGAATTCATCCCAGGAGAGGATTTCAGGCAAAGTGGTGTAATCCTCTTGGAAATGAAAT
5161 5220
CTCTTAAGTAGGGTCCTCTCCTAAAGTCCGTTTCACCACATTAGGAGAACCTTTACTTTA
e S F E D W S L I E P L T T Y D E Q F H F -
TGCTTGAGCTTACGATAGACGGTAGAGGTAGAGGTAGAGATGGCTAATTTAGAAGCGATA
5221 5280
ACGAACTCGAATGCTATCTGCCATCTCCATCTCCATCTCTACCGATTAAATCTTCGCTAT
e Q K L K R Y V T S T S T S I A L K S A I -
TGTGTAAGAGCCTCTCTGTTGAGTAGGAGTTGGGCAATTTTCTGTCTCACCATTTCCGAG
5281 5340
ACACATTCTCGGAGAGACAACTCATCCTCAACCCGTTAAAAGACAGAGTGGTAAAGGCTC
e H T L A E R N L L L Q A I K Q R V M E S -
ATTTGGCAATTTTTCTGAACGAGAGTTGTTTCAGCTACAGTGACTTTCCGACAGGACTTG
5341 5400
TAAACCGTTAAAAAGACTTGCTCTCAACAAAGTCGATGTCACTGAAAGGCTGTCCTGAAC
e I Q C N K Q V L T T E A V T V K R C S K -
CATTGAAATCGTCTCTTTTTCAAATGAATGAGGCTAGGGAAACCACCAATCTCGATAAAA
5401 5460
GTAACTTTAGCAGAGAAAAAGTTTACTTACTCCGATCCCTTTGGTGGTTAGAGCTATTTT
e C Q F R R K K L H I L S P F G G I E I F -
GGGATTTTAGAAGGCTTTTGGAAGTCGTATTTGATTTGTTTTCCTTTACAGTGTTTACAT
5461 5520
CCCTAAAATCTTCCGAAAACCTTCAGCATAAACTAAACAAAAGGAAATGTCACAAATGTA
e P I K S P K Q F D Y K I Q K G K C H K C -

Figure 101M

```
     TTAGGTGGGTGATAATCAAGTGTAGCGAAGACTTCGATATGGGTATCGTGCTGAATGGCT
5521 ---------+---------+---------+---------+---------+---------+ 5580
     AATCCACCCACTATTAGTTCACATCGCTTCTGAAGCTATACCCATAGCACGACTTACCGA

e     K  P  P  H  Y  D  L  T  A  F  V  E  I  H  T  D  H  Q  I  A  -

     TTATTTAAGGTGATGTTTTTGTCTTTTATTCCGATGAGTAATGTGGTATGATTGATGTGT
5581 ---------+---------+---------+---------+---------+---------+ 5640
     AATAAATTCCACTACAAAAACAGAAAATAAGGCTACTCATTACACCATACTAACTACACA

e     K  N  L  T  I  N  K  D  K  I  G  I  L  L  T  T  H  N  I  H  -

     TCCATAAGATACTTTCTAATGAGTTGTTTAGGCGCTTTTCATTATAAGTCTTATGGGACT
5641 ---------+---------+---------+---------+---------+---------+ 5700
     AGGTATTCTATGAAAGATTACTCAACAAATCCGCGAAAAGTAATATTCAGAATACCCTGA

e     E  M                                                       -

     TTTTTGATACTCAAAAAGCCCTATAATCTCCACAGTGGGATTTACCCACTACAGAAATTA
5701 ---------+---------+---------+---------+---------+---------+ 5760
     AAAAACTATGAGTTTTTCGGGATATTAGAGGTGTCACCCTAAATGGGTGATGTCTTTAAT


     TAGAGCCAGAAAAAACACTTTTGTTCACTAGCAGAAACTAGAGAGCAGAAGTGTTTTTCT
5761 ---------+---------+---------+---------+---------+---------+ 5820
     ATCTCGGTCTTTTTTGTGAAAACAAGTGATCGTCTTTGATCTCTCGTCTTCACAAAAAGA


     GTTCAGATTTACCCAAAACTGGGAAATATGGGGATAAGAATAGAGATGGCTTAGGAAGCC
5821 ---------+---------+---------+---------+---------+---------+ 5880
     CAAGTCTAAATGGGTTTTGACCCTTTATACCCCTATTCTTATCTCTACCGAATCCTTCGG


     CCTTTTTGTGTGTAGACAGTACGATGAACTTATAACAAATAGTGAGCCTTTTTAGCAATC
5881 ---------+---------+---------+---------+---------+---------+ 5940
     GGAAAAACACACATCTGTCATGCTACTTGAATATTGTTTATCACTCGGAAAAATCGTTAG

e                                  *  L  L  Y  H  A  K  K  A  I  -orf2_670 homologue of sp0461, transcriptional
                                                                     regulator
     ATTGCGACCCGTTTGTCAAAAGCCTCTTTTCGGATATCTACAATTGTCTGATAGATGAGA
```

Figure 101N

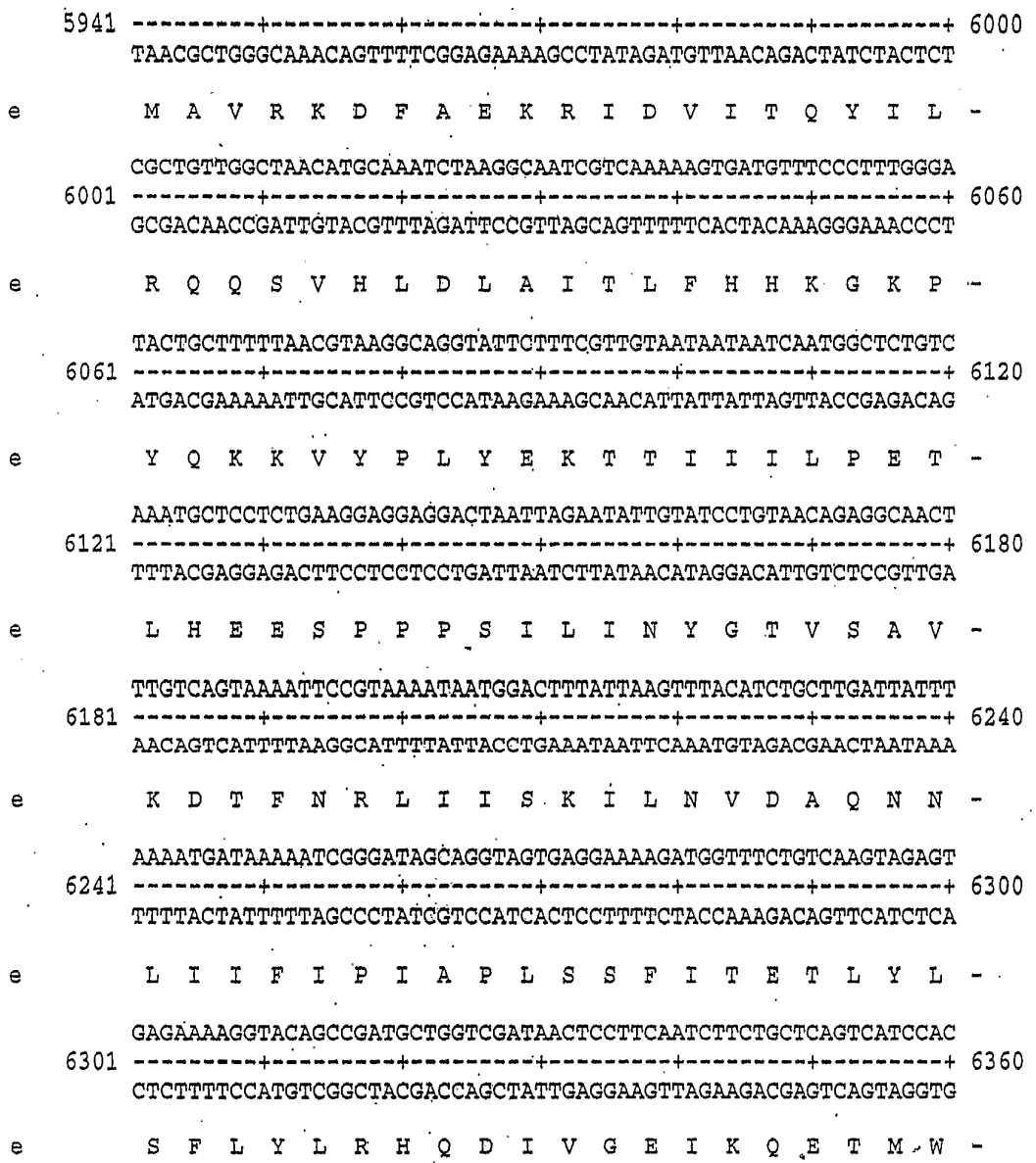

5941
6000
TAACGCTGGGCAAACAGTTTTCGGAGAAAAGCCTATAGATGTTAACAGACTATCTACTCT
e M A V R K D F A E K R I D V I T Q Y I L -
CGCTGTTGGCTAACATGCAAATCTAAGGCAATCGTCAAAAAGTGATGTTTCCCTTTGGGA
6001
6060
GCGACAACCGATTGTACGTTTAGATTCCGTTAGCAGTTTTTCACTACAAAGGGAAACCCT
e R Q Q S V H L D L A I T L F H H K G K P -
TACTGCTTTTTAACGTAAGGCAGGTATTCTTTCGTTGTAATAATAATCAATGGCTCTGTC
6061
6120
ATGACGAAAAATTGCATTCCGTCCATAAGAAAGCAACATTATTATTAGTTACCGAGACAG
e Y Q K K V Y P L Y E K T T I I I L P E T -
AAATGCTCCTCTGAAGGAGGAGGACTAATTAGAATATTGTATCCTGTAACAGAGGCAACT
6121
6180
TTTACGAGGAGACTTCCTCCTCCTGATTAATCTTATAACATAGGACATTGTCTCCGTTGA
e L H E E S P P P S I L I N Y G T V S A V -
TTGTCAGTAAAATTCCGTAAAATAATGGACTTTATTAAGTTTACATCTGCTTGATTATTT
6181
6240
AACAGTCATTTTAAGGCATTTTATTACCTGAAATAATTCAAATGTAGACGAACTAATAAA
e K D T F N R L I I S K I L N V D A Q N N -
AAAATGATAAAAATCGGGATAGCAGGTAGTGAGGAAAAGATGGTTTCTGTCAAGTAGAGT
6241
6300
TTTTACTATTTTTAGCCCTATCGTCCATCACTCCTTTTCTACCAAAGACAGTTCATCTCA
e L I I F I P I A P L S S F I T E T L Y L -
GAGAAAAGGTACAGCCGATGCTGGTCGATAACTCCTTCAATCTTCTGCTCAGTCATCCAC
6301
6360
CTCTTTTCCATGTCGGCTACGACCAGCTATTGAGGAAGTTAGAAGACGAGTCAGTAGGTG
e S F L Y L R H Q D I V G E I K Q E T M W -

Figure 1010

```
      TCTTGAACAATTGCTTTCGAAATATGATACAGTGGCTTGTCGCTTTCAATCCCATAATGT
6361  ---------+---------+---------+---------+---------+---------+ 6420
      AGAACTTGTTAACGAAAGCTTTATACTATGTCACCGAACAGCGAAAGTTAGGGTATTACA

e      E  Q  V  I  A  K  S  I  H  Y  L  P  K  D  S  E  I  G  Y  H  -

      TCGTAATAATTATAATAGGGAACTAGATTTTGTAAACCAAACAAAAACGTTCTTGTTAAG
6421  ---------+---------+---------+---------+---------+---------+ 6480
      AGCATTATTAATATTATCCCTTGATCTAAAACATTTGGTTTGTTTTTGCAAGAACAATTC

e      E  Y  Y  N  Y  Y  P  V  L  N  Q  L  G  F  L  F  T  R  T  L  -

      AAAGTCAGTGCTGTTAAAAAAGAAAGAGAATTCGAAATGTCATTTCCTAAGATATTCTTG
6481  ---------+---------+---------+---------+---------+---------+ 6540
      TTTCAGTCACGACAATTTTTTCTTTCTCTTAAGCTTTACAGTAAAGGATTCTATAAGAAC

e      F  T  L  A  T  L  F  S  L  S  N  S  I  D  N  G  L  I  N  K  -

      AACTTGGATAGTAGATGCTTTCCTCTTGTATGCTGAAGAATCAGTTGAATAGTATGAGTC
6541  ---------+---------+---------+---------+---------+---------+ 6600
      TTGAACCTATCATCTACGAAAGGAGAACATACGACTTCTTAGTCAACTTATCATACTCAG

e      F  K  S  L  L  H  K  G  R  T  H  Q  L  I  L  Q  I  T  H  T  -

      TTTTTTTCTTGATTCCATTTGTCCTTGGAAAACGAAGAATTAGCAGAACAATAAACCAAA
6601  ---------+---------+---------+---------+---------+---------+ 6660
      AAAAAAAGAACTAAGGTAAACAGGAACCTTTTGCTTCTTAATCGTCTTGTTATTTGGTTT

e      K  K  E  Q  N  W  K  D  K  S  F  S  S  N  A  S  C  Y  V  L  -

      AAGATATAATCCAGTTCTTCCTGAGTAAAAGTCATGTTGGCATGTGGCTCTAAGTAAGTT
6661  ---------+---------+---------+---------+---------+---------+ 6720
      TTCTATATTAGGTCAAGAAGGACTCATTTTCAGTACAACCGTACACCGAGATTCATTCAA

e      F  I  Y  D  L  E  E  Q  T  F  T  M  N  A  H  P  E  L  Y  T  -

      TGGCAATGTTCCATCAAAATCGGATACATAAAGAGGTTTTTTAATTTTTCAAACTCTTTG
6721  ---------+---------+---------+---------+---------+---------+ 6780
      ACCGTTACAAGGTAGTTTTAGCCTATGTATTTCTCCAAAAAATTAAAAAGTTTGAGAAAC
```

Figure 101P

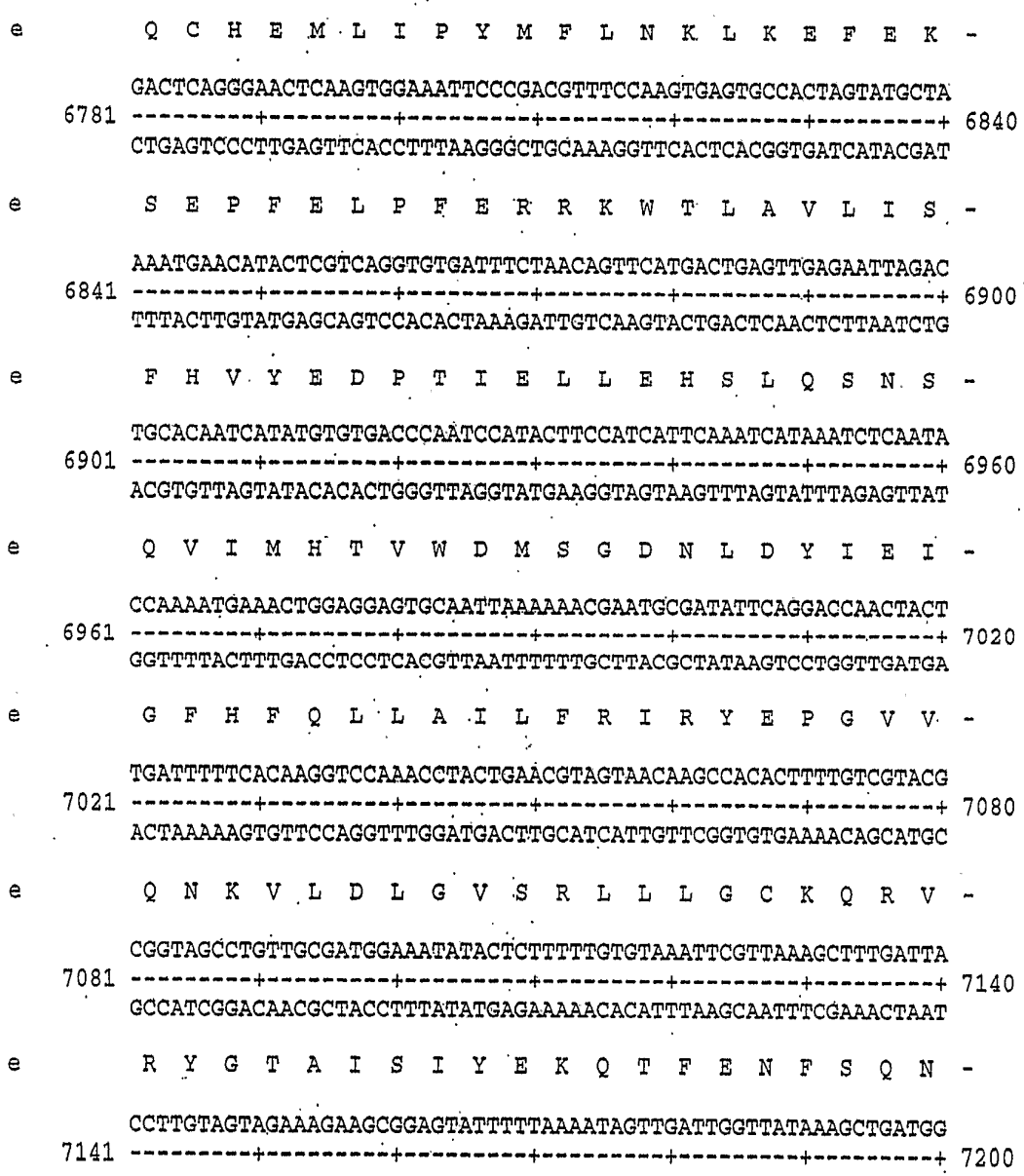

e Q C H E M L I P Y M F L N K L K E F E K -
GACTCAGGGAACTCAAGTGGAAATTCCCGACGTTTCCAAGTGAGTGCCACTAGTATGCTA
6781 ---------+---------+---------+---------+---------+---------+ 6840
CTGAGTCCCTTGAGTTCACCTTTAAGGGCTGCAAAGGTTCACTCACGGTGATCATACGAT
e S E P F E L P F E R R K W T L A V L I S -
AAATGAACATACTCGTCAGGTGTGATTTCTAACAGTTCATGACTGAGTTGAGAATTAGAC
6841 ---------+---------+---------+---------+---------+---------+ 6900
TTTACTTGTATGAGCAGTCCACACTAAAGATTGTCAAGTACTGACTCAACTCTTAATCTG
e F H V Y E D P T I E L L E H S L Q S N S -
TGCACAATCATATGTGTGACCCAATCCATACTTCCATCATTCAAATCATAAATCTCAATA
6901 ---------+---------+---------+---------+---------+---------+ 6960
ACGTGTTAGTATACACACTGGGTTAGGTATGAAGGTAGTAAGTTTAGTATTTAGAGTTAT
e Q V I M H T V W D M S G D N L D Y I E I -
CCAAAATGAAACTGGAGGAGTGCAATTAAAAAACGAATGCGATATTCAGGACCAACTACT
6961 ---------+---------+---------+---------+---------+---------+ 7020
GGTTTTACTTTGACCTCCTCACGTTAATTTTTTGCTTACGCTATAAGTCCTGGTTGATGA
e G F H F Q L L A I L F R I R Y E P G V V -
TGATTTTTCACAAGGTCCAAACCTACTGAACGTAGTAACAAGCCACACTTTTGTCGTACG
7021 ---------+---------+---------+---------+---------+---------+ 7080
ACTAAAAAGTGTTCCAGGTTTGGATGACTTGCATCATTGTTCGGTGTGAAAACAGCATGC
e Q N K V L D L G V S R L L L G C K Q R V -
CGGTAGCCTGTTGCGATGGAAATATACTCTTTTTGTGTAAATTCGTTAAAGCTTTGATTA
7081 ---------+---------+---------+---------+---------+---------+ 7140
GCCATCGGACAACGCTACCTTTATATGAGAAAAACACATTTAAGCAATTTCGAAACTAAT
e R Y G T A I S I Y E K Q T F E N F S Q N -
CCTTGTAGTAGAAAGAAGCGGAGTATTTTTAAAATAGTTGATTGGTTATAAAGCTGATGG
7141 ---------+---------+---------+---------+---------+---------+ 7200

Figure 101Q

```
          GGAACATCATCTTTCTTCGCCTCATAAAAATTTTATCAACTAACCAATATTTCGACTACC

e          G  Q  L  L  F  F  R  L  I  K  L  I  T  S  Q  N  Y  L  Q  H  -

          AAGTAATAATTCGTTTGATGAGAATGGTGTTCGATTAATTGAACTTGTTGCGTATCTAAA
     7201 ---------+---------+---------+---------+---------+---------+ 7260
          TTCATTATTAAGCAAACTACTCTTACCACAAGCTAATTAACTTGAACAACGCATAGATTT

e          F  Y  Y  N  T  Q  H  S  H  H  E  I  L  Q  V  Q  Q  T  D  L  -

          TTAAATGTCAACTCTTCCTCGAATGTTTCTTGTAATTCCTGCAAAATGCTTAGGAGACTT
     7261 ---------+---------+---------+---------+---------+---------+ 7320
          AATTTACAGTTGAGAAGGAGCTTACAAAGAACATTAAGGACGTTTTACGAATCCTCTGAA

e          N  F  T  L  E  E  E  F  T  E  Q  L  E  Q  L  I  S  L  L  S  -

          TTAGATTGTAATGAAGTTAAAGTAGACAGTTCATCTAGTTCAATAGACCGAATATCCAAT
     7321 ---------+---------+---------+---------+---------+---------+ 7380
          AATCTAACATTACTTCAATTTCATCTGTCAAGTAGATCAAGTTATCTGGCTTATAGGTTA

e          K  S  Q  L  S  T  L  T  S  L  E  D  L  E  I  S  R  I  D  L  -

          AATATATTTAAAATGGTAATTTTATCTGTAATTCTTTTTTCAATGTATTTGTTTAGCATA
     7381 ---------+---------+---------+---------+---------+---------+ 7440
          TTATATAAATTTTACCATTAAAATAGACATTAAGAAAAAAGTTACATAAACAAATCGTAT

e          L  I  N  L  I  T  I  K  D  T  I  R  K  E  I  Y  K  N  L  M  -

          GTTACCGAATCTTAGTTGCATATAGATAATTTTAATTATTATAATACAAAAGAAACTAAT
     7441 ---------+---------+---------+---------+---------+---------+ 7500
          CAATGGCTTAGAATCAACGTATATCTATTAAAATTAATAATATTATGTTTTCTTTGATTA


          TGTCTTGTCAAAAAGGTTGTGGAATTTCCGACTTTATTGATAAAACAGCATGTAATAAAA
     7501 ---------+---------+---------+---------+---------+---------+ 7560
          ACAGAACAGTTTTTCCAACACCTTAAAGGCTGAAATAACTATTTTGTCGTACATTATTTT


          GGCATTTTAAAGATAGTAATGAGTATTGGTGGAGTTTTATGGCTTATTTTTTTTATTAGA
     7561 ---------+---------+---------+---------+---------+---------+ 7620
```

Figure 101R

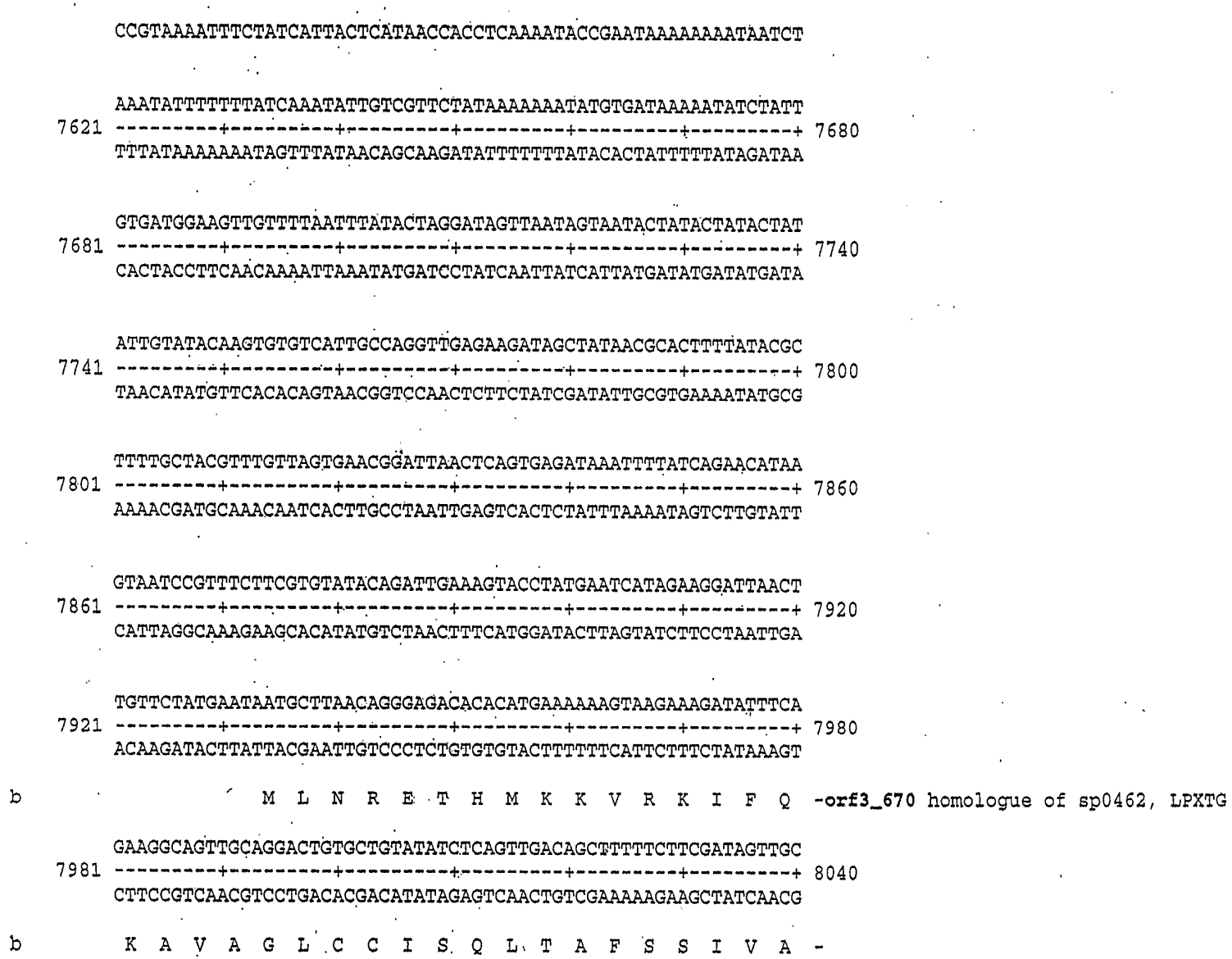

CCGTAAAATTTCTATCATTACTCATAACCACCTCAAAATACCGAATAAAAAAAATAATCT
AAATATTTTTTTATCAAATATTGTCGTTCTATAAAAAAATATGTGATAAAAATATCTATT
7621 ---------+---------+---------+---------+---------+---------+ 7680
TTTATAAAAAAATAGTTTATAACAGCAAGATATTTTTTTATACACTATTTTTATAGATAA
GTGATGGAAGTTGTTTTAATTTATACTAGGATAGTTAATAGTAATACTATACTATACTAT
7681 ---------+---------+---------+---------+---------+---------+ 7740
CACTACCTTCAACAAAATTAAATATGATCCTATCAATTATCATTATGATATGATATGATA
ATTGTATACAAGTGTGTCATTGCCAGGTTGAGAAGATAGCTATAACGCACTTTTATACGC
7741 ---------+---------+---------+---------+---------+---------+ 7800
TAACATATGTTCACACAGTAACGGTCCAACTCTTCTATCGATATTGCGTGAAAATATGCG
TTTTGCTACGTTTGTTAGTGAACGGATTAACTCAGTGAGATAAATTTTATCAGAACATAA
7801 ---------+---------+---------+---------+---------+---------+ 7860
AAAACGATGCAAACAATCACTTGCCTAATTGAGTCACTCTATTTAAAATAGTCTTGTATT
GTAATCCGTTTCTTCGTGTATACAGATTGAAAGTACCTATGAATCATAGAAGGATTAACT
7861 ---------+---------+---------+---------+---------+---------+ 7920
CATTAGGCAAAGAAGCACATATGTCTAACTTTCATGGATACTTAGTATCTTCCTAATTGA
TGTTCTATGAATAATGCTTAACAGGGAGACACACATGAAAAAAGTAAGAAAGATATTTCA
7921 ---------+---------+---------+---------+---------+---------+ 7980
ACAAGATACTTATTACGAATTGTCCCTCTGTGTGTACTTTTTTCATTCTTTCTATAAAGT
b M L N R E T H M K K V R K I F Q -orf3_670 homologue of sp0462, LPXTG
GAAGGCAGTTGCAGGACTGTGCTGTATATCTCAGTTGACAGCTTTTTCTTCGATAGTTGC
7981 ---------+---------+---------+---------+---------+---------+ 8040
CTTCCGTCAACGTCCTGACACGACATATAGAGTCAACTGTCGAAAAAGAAGCTATCAACG
b K A V A G L C C I S Q L T A F S S I V A -

Figure 101S

```
     TTTAGCAGAAACGCCTGAAACCAGTCCAGCGATAGGAAAAGTAGTGATTAAGGAGACAGG
8041 ---------+---------+---------+---------+---------+---------+ 8100
     AAATCGTCTTTGCGGACTTTGGTCAGGTCGCTATCCTTTTCATCACTAATTCCTCTGTCC

b      L  A  E  T  P  E  T  S  P  A  I  G  K  V  V  I  K  E  T  G  -

     CGAAGGAGGAGCGCTTCTAGGAGATGCCGTCTTTGAGTTGAAAAACAATACGGATGGCAC
8101 ---------+---------+---------+---------+---------+---------+ 8160
     GCTTCCTCCTCGCGAAGATCCTCTACGGCAGAAACTCAACTTTTTGTTATGCCTACCGTG

b      E  G  G  A  L  L  G  D  A  V  F  E  L  K  N  N  T  D  G  T  -

     AACTGTTTCGCAAAGGACAGAGGCGCAAACAGGAGAAGCGATATTTTCAAACATAAAACC
8161 ---------+---------+---------+---------+---------+---------+ 8220
     TTGACAAAGCGTTTCCTGTCTCCGCGTTTGTCCTCTTCGCTATAAAAGTTTGTATTTTGG

b      T  V  S  Q  R  T  E  A  Q  T  G  E  A  I  F  S  N  I  K  P  -

     TGGGACATACACCTTGACAGAAGCCCAACCTCCAGTTGGTTATAAACCCTCTACTAAACA
8221 ---------+---------+---------+---------+---------+---------+ 8280
     ACCCTGTATGTGGAACTGTCTTCGGGTTGGAGGTCAACCAATATTTGGGAGATGATTTGT

b      G  T  Y  T  L  T  E  A  Q  P  P  V  G  Y  K  P  S  T  K  Q  -

     ATGGACTGTTGAAGTTGAGAAGAATGGTCGGACGACTGTCCAAGGTGAACAGGTAGAAAA
8281 ---------+---------+---------+---------+---------+---------+ 8340
     TACCTGACAACTTCAACTCTTCTTACCAGCCTGCTGACAGGTTCCACTTGTCCATCTTTT

b      W  T  V  E  V  E  K  N  G  R  T  T  V  Q  G  E  Q  V  E  N  -

     TCGAGAAGAGGCTCTATCTGACCAGTATCCACAAACAGGGACTTATCCAGATGTTCAAAC
8341 ---------+---------+---------+---------+---------+---------+ 8400
     AGCTCTTCTCCGAGATAGACTGGTCATAGGTGTTTGTCCCTGAATAGGTCTACAAGTTTG

b      R  E  E  A  L  S  D  Q  Y  P  Q  T  G  T  Y  P  D  V  Q  T  -

     ACCTTATCAGATTATTAAGGTAGATGGTTCGGAAAAAAACGGACAGCACAAGGCGTTGAA
8401 ---------+---------+---------+---------+---------+---------+ 8460
     TGGAATAGTCTAATAATTCCATCTACCAAGCCTTTTTTTGCCTGTCGTGTTCCGCAACTT
```

Figure 101T

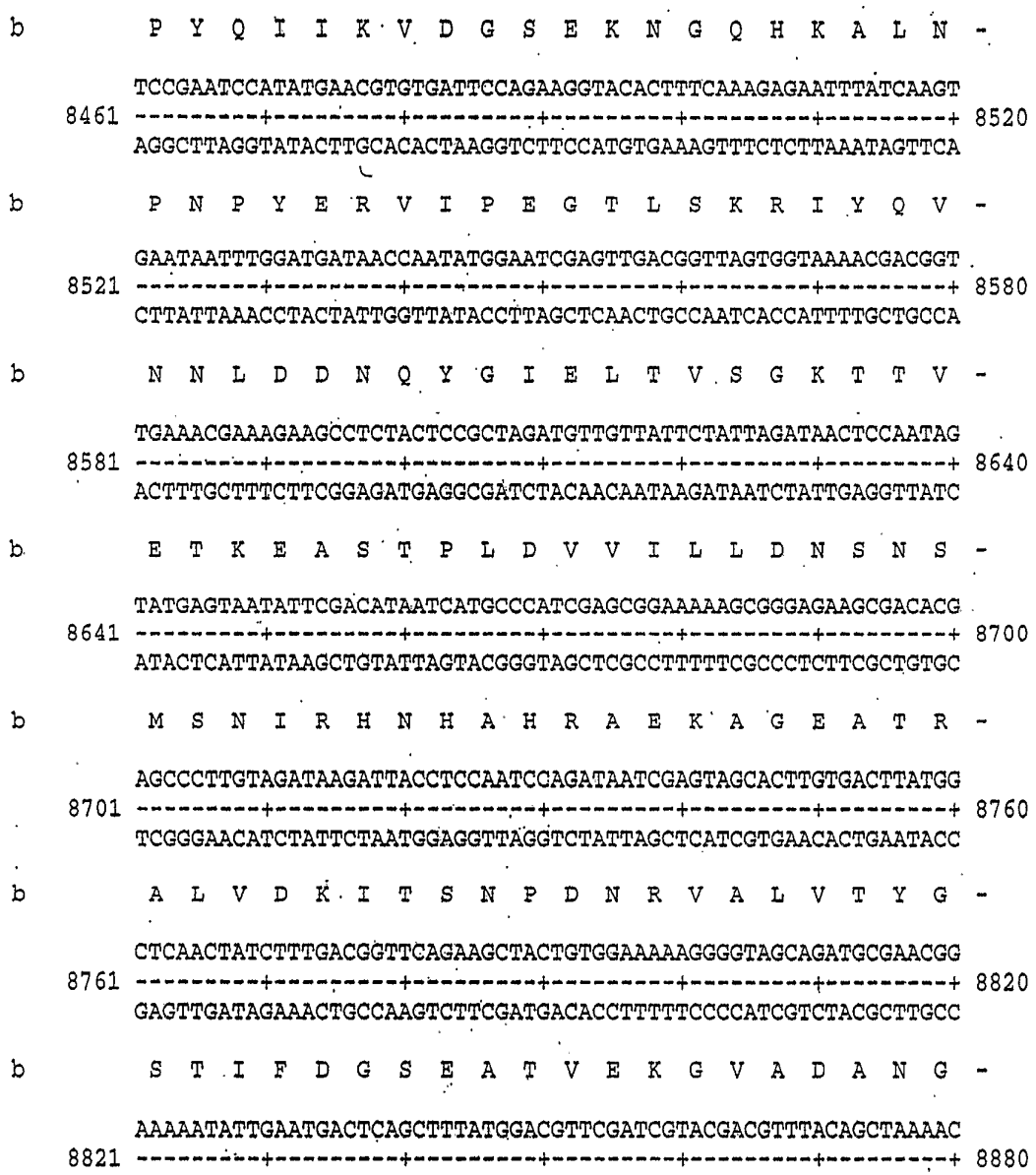

b P Y Q I I K V D G S E K N G Q H K A L N -
TCCGAATCCATATGAACGTGTGATTCCAGAAGGTACACTTTCAAAGAGAATTTATCAAGT
8461 ---------+---------+---------+---------+---------+---------+ 8520
AGGCTTAGGTATACTTGCACACTAAGGTCTTCCATGTGAAAGTTTCTCTTAAATAGTTCA
b P N P Y E R V I P E G T L S K R I Y Q V -
GAATAATTTGGATGATAACCAATATGGAATCGAGTTGACGGTTAGTGGTAAAACGACGGT
8521 ---------+---------+---------+---------+---------+---------+ 8580
CTTATTAAACCTACTATTGGTTATACCTTAGCTCAACTGCCAATCACCATTTTGCTGCCA
b N N L D D N Q Y G I E L T V S G K T T V -
TGAAACGAAAGAAGCCTCTACTCCGCTAGATGTTGTTATTCTATTAGATAACTCCAATAG
8581 ---------+---------+---------+---------+---------+---------+ 8640
ACTTTGCTTTCTTCGGAGATGAGGCGATCTACAACAATAAGATAATCTATTGAGGTTATC
b E T K E A S T P L D V V I L L D N S N S -
TATGAGTAATATTCGACATAATCATGCCCATCGAGCGGAAAAAGCGGGAGAAGCGACACG
8641 ---------+---------+---------+---------+---------+---------+ 8700
ATACTCATTATAAGCTGTATTAGTACGGGTAGCTCGCCTTTTTCGCCCTCTTCGCTGTGC
b M S N I R H N H A H R A E K A G E A T R -
AGCCCTTGTAGATAAGATTACCTCCAATCCAGATAATCGAGTAGCACTTGTGACTTATGG
8701 ---------+---------+---------+---------+---------+---------+ 8760
TCGGGAACATCTATTCTAATGGAGGTTAGGTCTATTAGCTCATCGTGAACACTGAATACC
b A L V D K I T S N P D N R V A L V T Y G -
CTCAACTATCTTTGACGGTTCAGAAGCTACTGTGGAAAAAGGGGTAGCAGATGCGAACGG
8761 ---------+---------+---------+---------+---------+---------+ 8820
GAGTTGATAGAAACTGCCAAGTCTTCGATGACACCTTTTTCCCCATCGTCTACGCTTGCC
b S T I F D G S E A T V E K G V A D A N G -
AAAAATATTGAATGACTCAGCTTTATGGACGTTCGATCGTACGACGTTTACAGCTAAAAC
8821 ---------+---------+---------+---------+---------+---------+ 8880

Figure 101U

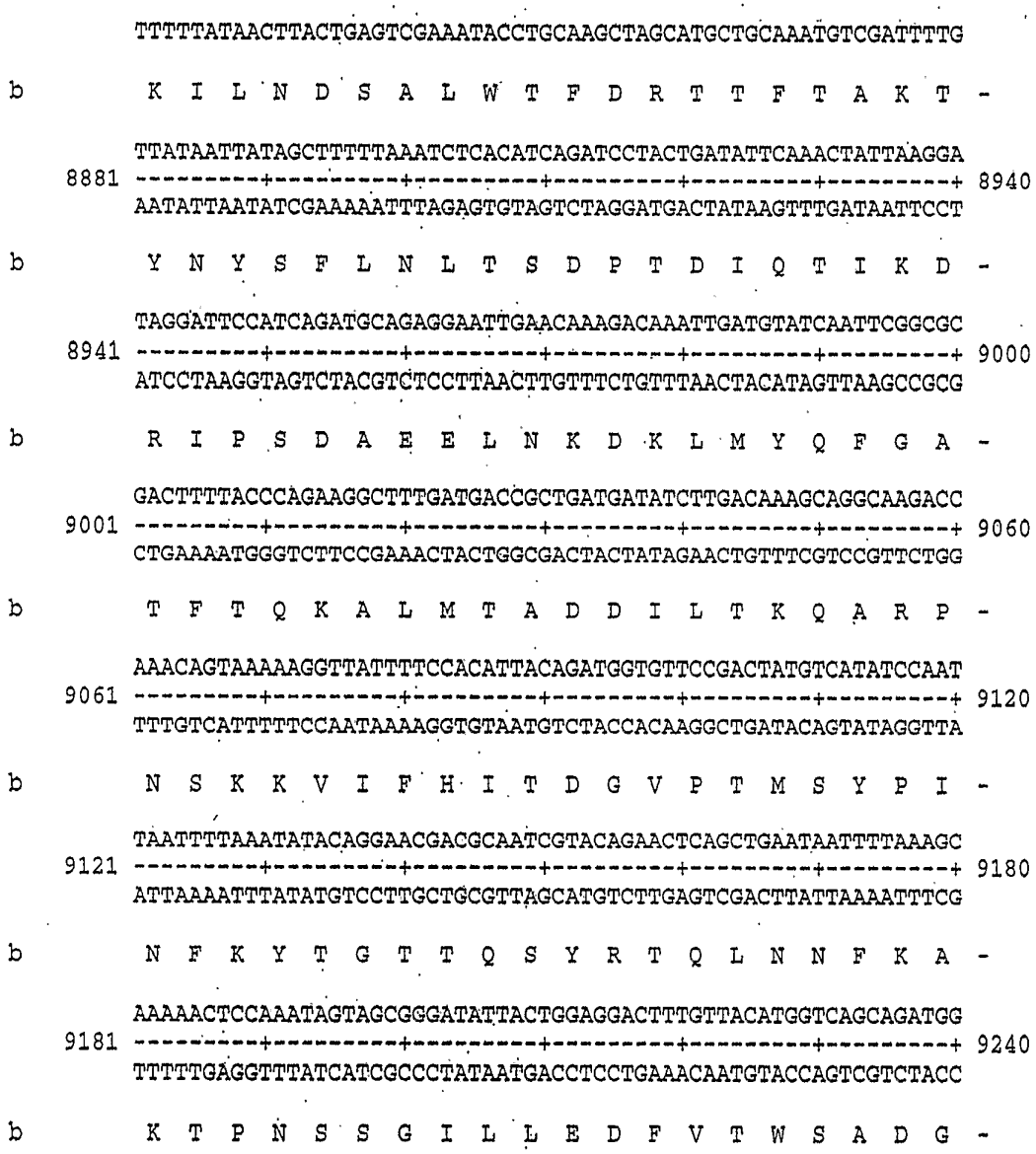

TTTTTATAACTTACTGAGTCGAAATACCTGCAAGCTAGCATGCTGCAAATGTCGATTTTG
b K I L N D S A L W T F D R T T F T A K T -
TTATAATTATAGCTTTTTAAATCTCACATCAGATCCTACTGATATTCAAACTATTAAGGA
8881 + 8940
AATATTAATATCGAAAAATTTAGAGTGTAGTCTAGGATGACTATAAGTTTGATAATTCCT
b Y N Y S F L N L T S D P T D I Q T I K D -
TAGGATTCCATCAGATGCAGAGGAATTGAACAAAGACAAATTGATGTATCAATTCGGCGC
8941 + 9000
ATCCTAAGGTAGTCTACGTCTCCTTAACTTGTTTCTGTTTAACTACATAGTTAAGCCGCG
b R I P S D A E E L N K D K L M Y Q F G A -
GACTTTTACCCAGAAGGCTTTGATGACCGCTGATGATATCTTGACAAAGCAGGCAAGACC
9001 + 9060
CTGAAAATGGGTCTTCCGAAACTACTGGCGACTACTATAGAACTGTTTCGTCCGTTCTGG
b T F T Q K A L M T A D D I L T K Q A R P -
AAACAGTAAAAAGGTTATTTTCCACATTACAGATGGTGTTCCGACTATGTCATATCCAAT
9061 + 9120
TTTGTCATTTTTCCAATAAAAGGTGTAATGTCTACCACAAGGCTGATACAGTATAGGTTA
b N S K K V I F H I T D G V P T M S Y P I -
TAATTTTAAATATACAGGAACGACGCAATCGTACAGAACTCAGCTGAATAATTTTAAAGC
9121 + 9180
ATTAAAATTTATATGTCCTTGCTGCGTTAGCATGTCTTGAGTCGACTTATTAAAATTTCG
b N F K Y T G T T Q S Y R T Q L N N F K A -
AAAAACTCCAAATAGTAGCGGGATATTACTGGAGGACTTTGTTACATGGTCAGCAGATGG
9181 + 9240
TTTTTGAGGTTTATCATCGCCCTATAATGACCTCCTGAAACAATGTACCAGTCGTCTACC
b K T P N S S G I L L E D F V T W S A D G -

Figure 101V

```
      TGAACATAAGATTGTTCGTGGAGATGGTGAAAGTTATCAGATGTTTACGAAGAAACCTGT
 9241 ---------+---------+---------+---------+---------+---------+ 9300
      ACTTGTATTCTAACAAGCACCTCTACCACTTTCAATAGTCTACAAATGCTTCTTTGGACA

b      E  H  K  I  V  R  G  D  G  E  S  Y  Q  M  F  T  K  K  P  V  -

      AACAGACCAATACGGAGTTCATCAAATACTTTCAATCACCTCCATGGAGCAGAGAGCTAA
 9301 ---------+---------+---------+---------+---------+---------+ 9360
      TTGTCTGGTTATGCCTCAAGTAGTTTATGAAAGTTAGTGGAGGTACCTCGTCTCTCGATT

b      T  D  Q  Y  G  V  H  Q  I  L  S  I  T  S  M  E  Q  R  A  K  -

      ATTAGTTTCAGCGGGATATAGGTTCTATGGAACTGACTTGTATTTATATTGGCGTGATAG
 9361 ---------+---------+---------+---------+---------+---------+ 9420
      TAATCAAAGTCGCCCTATATCCAAGATACCTTGACTGAACATAAATATAACCGCACTATC

b      L  V  S  A  G  Y  R  F  Y  G  T  D  L  Y  L  Y  W  R  D  S  -

      TATTCTAGCCTATCCATTTAACTCTAGTACCGATTGGATTACCAACCATGGTGACCCTAC
 9421 ---------+---------+---------+---------+---------+---------+ 9480
      ATAAGATCGGATAGGTAAATTGAGATCATGGCTAACCTAATGGTTGGTACCACTGGGATG

b      I  L  A  Y  P  F  N  S  S  T  D  W  I  T  N  H  G  D  P  T  -

      GACTTGGTATTATAACGGAAATATGGCTCAGGATGGCTATGATGTCTTCACTGTTGGGGT
 9481 ---------+---------+---------+---------+---------+---------+ 9540
      CTGAACCATAATATTGCCTTTATACCGAGTCCTACCGATACTACAGAAGTGACAACCCCA

b      T  W  Y  Y  N  G  N  M  A  Q  D  G  Y  D  V  F  T  V  G  V  -

      TGGTGTAAACGGGGATCCTGGTACGGATGAAGCAACGGCTACTAGATTTATGCAGAGCAT
 9541 ---------+---------+---------+---------+---------+---------+ 9600
      ACCACATTTGCCCCTAGGACCATGCCTACTTCGTTGCCGATGATCTAAATACGTCTCGTA

b      G  V  N  G  D  P  G  T  D  E  A  T  A  T  R  F  M  Q  S  I  -

      CTCTAGTTCTCCTGACAACTACACTAACGTAGCAGATCCATCTCAGATTTTACAAGAATT
 9601 ---------+---------+---------+---------+---------+---------+ 9660
      GAGATCAAGAGGACTGTTGATGTGATTGCATCGTCTAGGTAGAGTCTAAAATGTTCTTAA
```

Figure 101W

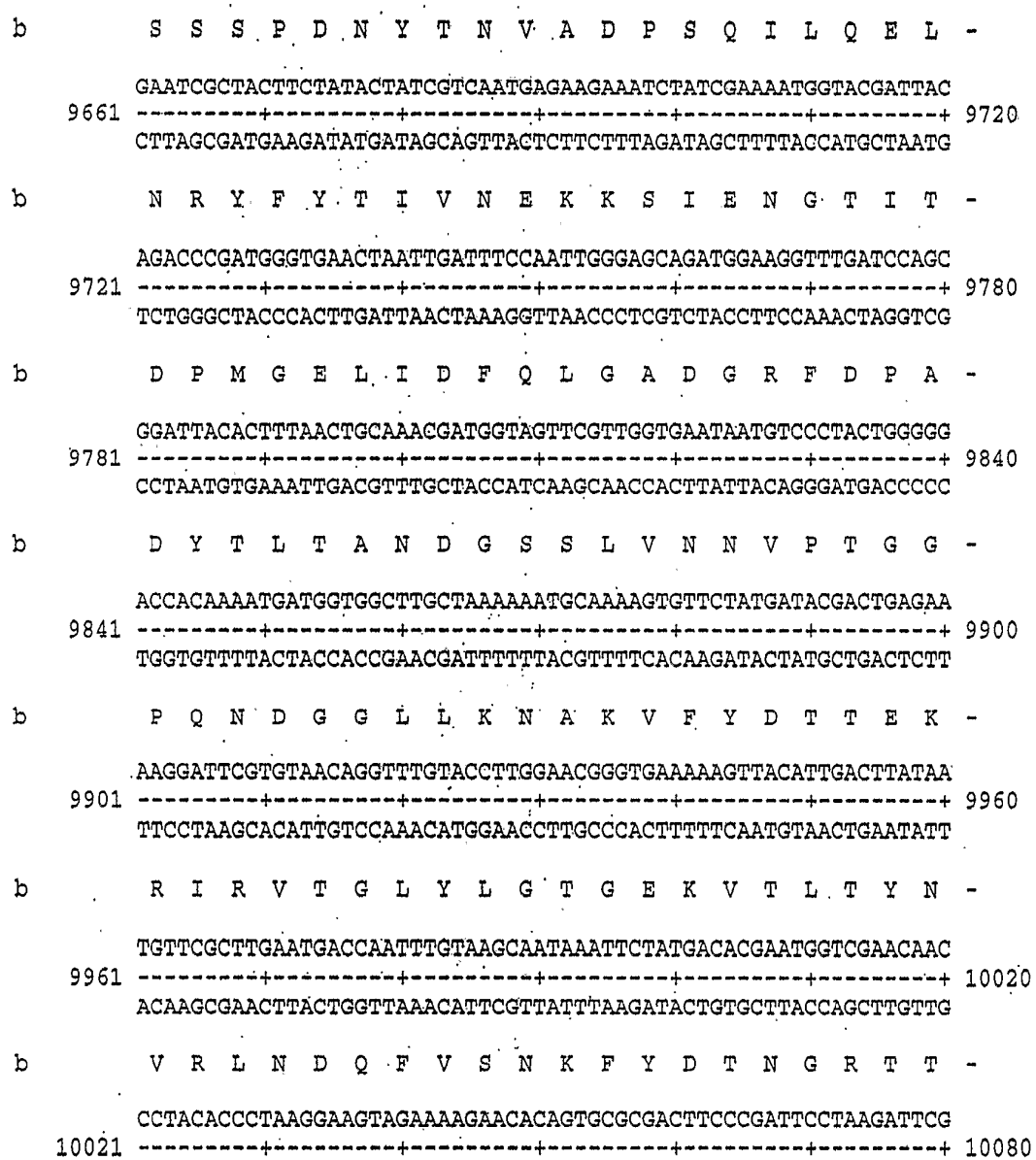

b S S S P D N Y T N V A D P S Q I L Q E L -
GAATCGCTACTTCTATACTATCGTCAATGAGAAGAAATCTATCGAAAATGGTACGATTAC
9661 + 9720
CTTAGCGATGAAGATATGATAGCAGTTACTCTTCTTTAGATAGCTTTTACCATGCTAATG
b N R Y F Y T I V N E K K S I E N G T I T -
AGACCCGATGGGTGAACTAATTGATTTCCAATTGGGAGCAGATGGAAGGTTTGATCCAGC
9721 + 9780
TCTGGGCTACCCACTTGATTAACTAAAGGTTAACCCTCGTCTACCTTCCAAACTAGGTCG
b D P M G E L I D F Q L G A D G R F D P A -
GGATTACACTTTAACTGCAAACGATGGTAGTTCGTTGGTGAATAATGTCCCTACTGGGGG
9781 + 9840
CCTAATGTGAAATTGACGTTTGCTACCATCAAGCAACCACTTATTACAGGGATGACCCCC
b D Y T L T A N D G S S L V N N V P T G G -
ACCACAAAATGATGGTGGCTTGCTAAAAAATGCAAAAGTGTTCTATGATACGACTGAGAA
9841 + 9900
TGGTGTTTTACTACCACCGAACGATTTTTTACGTTTTCACAAGATACTATGCTGACTCTT
b P Q N D G G L L K N A K V F Y D T T E K -
AAGGATTCGTGTAACAGGTTTGTACCTTGGAACGGGTGAAAAAGTTACATTGACTTATAA
9901 + 9960
TTCCTAAGCACATTGTCCAAACATGGAACCTTGCCCACTTTTTCAATGTAACTGAATATT
b R I R V T G L Y L G T G E K V T L T Y N -
TGTTCGCTTGAATGACCAATTTGTAAGCAATAAATTCTATGACACGAATGGTCGAACAAC
9961 + 10020
ACAAGCGAACTTACTGGTTAAACATTCGTTATTTAAGATACTGTGCTTACCAGCTTGTTG
b V R L N D Q F V S N K F Y D T N G R T T -
CCTACACCCTAAGGAAGTAGAAAAGAACACAGTGCGCGACTTCCCGATTCCTAAGATTCG
10021 + 10080

Figure 101X

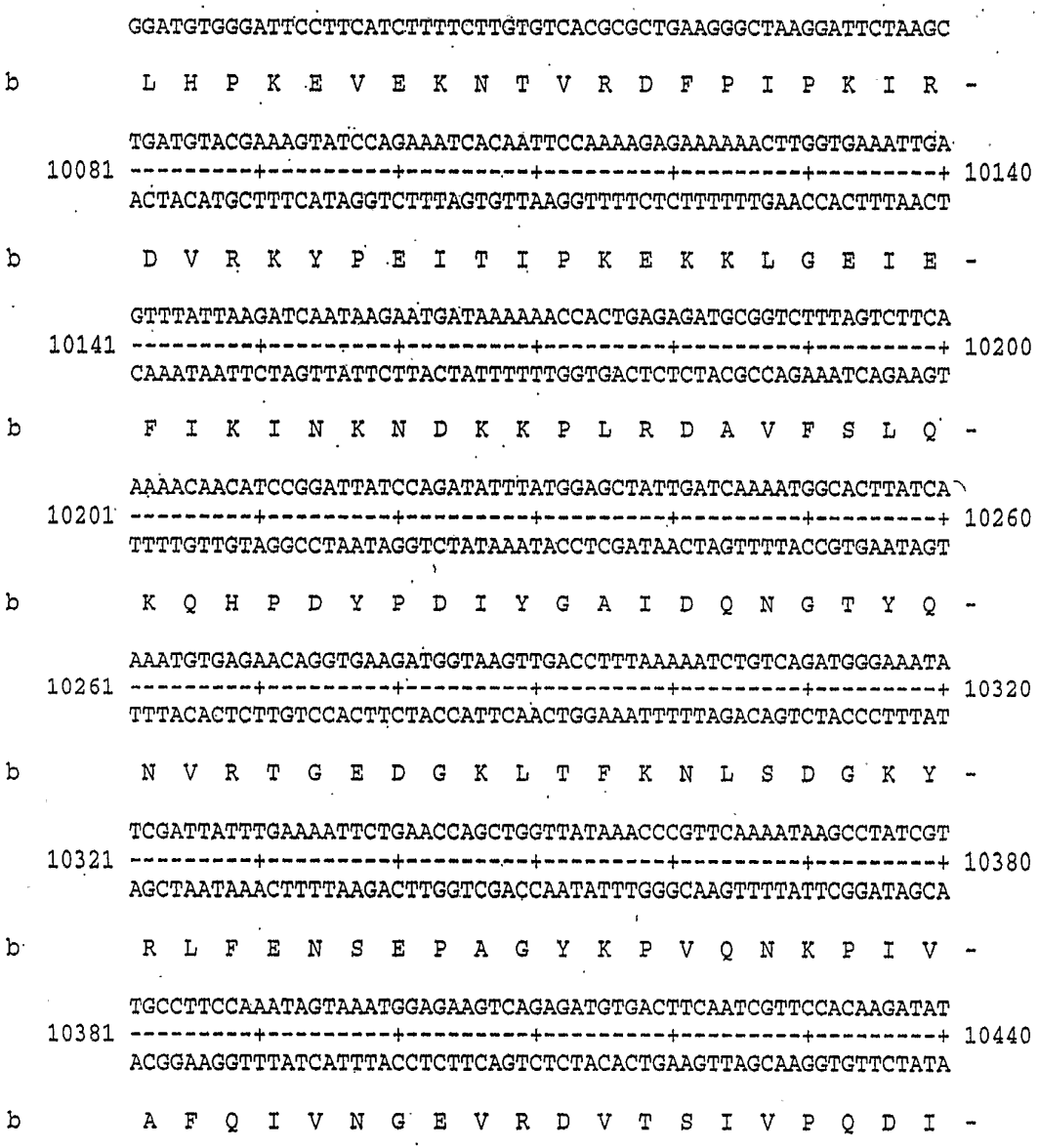

GGATGTGGGATTCCTTCATCTTTTCTTGTGTCACGCGCTGAAGGGCTAAGGATTCTAAGC
b L H P K E V E K N T V R D F P I P K I R -
TGATGTACGAAAGTATCCAGAAATCACAATTCCAAAAGAGAAAAAACTTGGTGAAATTGA
10081 ---------+---------+---------+---------+---------+---------+ 10140
ACTACATGCTTTCATAGGTCTTTAGTGTTAAGGTTTTCTCTTTTTTGAACCACTTTAACT
b D V R K Y P E I T I P K E K K L G E I E -
GTTTATTAAGATCAATAAGAATGATAAAAAACCACTGAGAGATGCGGTCTTTAGTCTTCA
10141 ---------+---------+---------+---------+---------+---------+ 10200
CAAATAATTCTAGTTATTCTTACTATTTTTTGGTGACTCTCTACGCCAGAAATCAGAAGT
b F I K I N K N D K K P L R D A V F S L Q -
AAAACAACATCCGGATTATCCAGATATTTATGGAGCTATTGATCAAAATGGCACTTATCA
10201 ---------+---------+---------+---------+---------+---------+ 10260
TTTTGTTGTAGGCCTAATAGGTCTATAAATACCTCGATAACTAGTTTTACCGTGAATAGT
b K Q H P D Y P D I Y G A I D Q N G T Y Q -
AAATGTGAGAACAGGTGAAGATGGTAAGTTGACCTTTAAAAATCTGTCAGATGGGAAATA
10261 ---------+---------+---------+---------+---------+---------+ 10320
TTTACACTCTTGTCCACTTCTACCATTCAACTGGAAATTTTTAGACAGTCTACCCTTTAT
b N V R T G E D G K L T F K N L S D G K Y -
TCGATTATTTGAAAATTCTGAACCAGCTGGTTATAAACCCGTTCAAAATAAGCCTATCGT
10321 ---------+---------+---------+---------+---------+---------+ 10380
AGCTAATAAACTTTTAAGACTTGGTCGACCAATATTTGGGCAAGTTTTATTCGGATAGCA
b R L F E N S E P A G Y K P V Q N K P I V -
TGCCTTCCAAATAGTAAATGGAGAAGTCAGAGATGTGACTTCAATCGTTCCACAAGATAT
10381 ---------+---------+---------+---------+---------+---------+ 10440
ACGGAAGGTTTATCATTTACCTCTTCAGTCTCTACACTGAAGTTAGCAAGGTGTTCTATA
b A F Q I V N G E V R D V T S I V P Q D I -

Figure 101Y

```
      ACCAGCGGGTTACGAGTTTACGAATGATAAGCACTATATCACAAATGAGCCAATTCCTCC
10441 ---------+---------+---------+---------+---------+---------+ 10500
      TGGTCGCCCAATGCTCAAATGCTTACTATTCGTGATATAGTGTTTACTCGGTTAAGGAGG

b       P  A  G  Y  E  F  T  N  D  K  H  Y  I  T  N  E  P  I  P  P  -

      AAAAAGAGAATATCCTCGAACTGGTGGTATCGGAATGTTGCCATTCTATCTGATAGGTTG
10501 ---------+---------+---------+---------+---------+---------+ 10560
      TTTTTCTCTTATAGGAGCTTGACCACCATAGCCTTACAACGGTAAGATAGACTATCCAAC

b       K  R  E  Y  P  R  T  G  G  I  G  M  L  P  F  Y  L  I  G  C  -

      CATGATGATGGGAGGAGTTCTATTATACACACGGAAACATCCGTAAAGTGTAGCAATGAG
10561 ---------+---------+---------+---------+---------+---------+ 10620
      GTACTACTACCCTCCTCAAGATAATATGTGTGCCTTTGTAGGCATTTCACATCGTTACTC

b       M  M  M  G  G  V  L  L  Y  T  R  K  H  P  *

      AAATGATAATATCGATACTCTGAGCGATACTTTTAAGAAGTAGCACTCAAGAAGAGATTT
10621 ---------+---------+---------+---------+---------+---------+ 10680
      TTTACTATTATAGCTATGAGACTCGCTATGAAAATTCTTCATCGTGAGTTCTTCTCTAAA

      AAGTTTACTTGGTGAAAACAGTTTTCTTCGCCAAGTAAACCACCATTGAAAGGGGAGATG
10681 ---------+---------+---------+---------+---------+---------+ 10740
      TTCAAATGAACCACTTTTGTCAAAAGAAGCGGTTCATTTGGTGGTAACTTTCCCCTCTAC

      TTTTCGAAAACTTGCACAGAAAAAAGGATTATTATTGTCATGTGTAATTCATTACATTGC
10741 ---------+---------+---------+---------+---------+---------+ 10800
      AAAAGCTTTTGAACGTGTCTTTTTTCCTAATAATAACAGTACACATTAAGTAATGTAACG

      TCACAGTTGATTTTAAGAGATATGAATAAGGAGAAATCATGAAATCAATCAACAAATTTT
10801 ---------+---------+---------+---------+---------+---------+ 10860
      AGTGTCAACTAAAATTCTCTATACTTATTCCTCTTTAGTACTTTAGTTAGTTGTTTAAAA

c                                                 M  K  S  I  N  K  F  L  - orf4_670, homologue of sp0463, LPXTG

      TAACAATGCTTGCTGCCTTATTACTGACAGCGAGTAGCCTGTTTTCAGCTGCAACAGTTT
```

Figure 101Z

```
 10861 ---------+---------+---------+---------+---------+---------+ 10920
       ATTGTTACGAACGACGGAATAATGACTGTCGCTCATCGGACAAAAGTCGACGTTGTCAAA

c        T  M  L  A  A  L  L  L  T  A  S  S  L  F  S  A  A  T  V  F -

       TTGCGGCGGACAATGTTAGTACAGCACCAGATGCTGTTACTAAAACTTTAACAATCCATA
 10921 ---------+---------+---------+---------+---------+---------+ 10980
       AACGCCGCCTGTTACAATCATGTCGTGGTCTACGACAATGATTTTGAAATTGTTAGGTAT

c        A  A  D  N  V  S  T  A  P  D  A  V  T  K  T  L  T  I  H  K -

       AGTTACTGCTCTCAGAAGATGATTTAAAGACTTGGGATACAAACGGTCCTAAAGGATATG
 10981 ---------+---------+---------+---------+---------+---------+ 11040
       TCAATGACGAGAGTCTTCTACTAAATTTCTGAACCCTATGTTTGCCAGGATTTCCTATAC

c        L  L  L  S  E  D  D  L  K  T  W  D  T  N  G  P  K  G  Y  D -

       ATGGAACTCAATCTAGTTTAAAAGATTTAACTGGAGTTGTAGCTGAGGAAATTCCAAATG
 11041 ---------+---------+---------+---------+---------+---------+ 11100
       TACCTTGAGTTAGATCAAATTTTCTAAATTGACCTCAACATCGACTCCTTTAAGGTTTAC

c        G  T  Q  S  S  L  K  D  L  T  G  V  V  A  E  E  I  P  N  V -

       TATACTTTGAATTACAAAAGTATAATTTGACTGATGGTAAGGAAAAAGAAAATCTTAAAG
 11101 ---------+---------+---------+---------+---------+---------+ 11160
       ATATGAAACTTAATGTTTTCATATTAAACTGACTACCATTCCTTTTTCTTTTAGAATTTC

c        Y  F  E  L  Q  K  Y  N  L  T  D  G  K  E  K  E  N  L  K  D -

       ATGATAGTAAATGGACAACAGTTCATGGTGGTTTGACAACTAAAGATGGACTTAAAATTG
 11161 ---------+---------+---------+---------+---------+---------+ 11220
       TACTATCATTTACCTGTTGTCAAGTACCACCAAACTGTTGATTTCTACCTGAATTTTAAC

c        D  S  K  W  T  T  V  H  G  G  L  T  T  K  D  G  L  K  I  E -

       AAACCAGTACTCTTAAAGGTGTGTATCGTATTCGTGAGGATAGAACAAAGACTACCTATG
 11221 ---------+---------+---------+---------+---------+---------+ 11280
       TTTGGTCATGAGAATTTCCACACATAGCATAAGCACTCCTATCTTGTTTCTGATGGATAC

c        T  S  T  L  K  G  V  Y  R  I  R  E  D  R  T  K  T  T  Y  V -
```

Figure 101AA

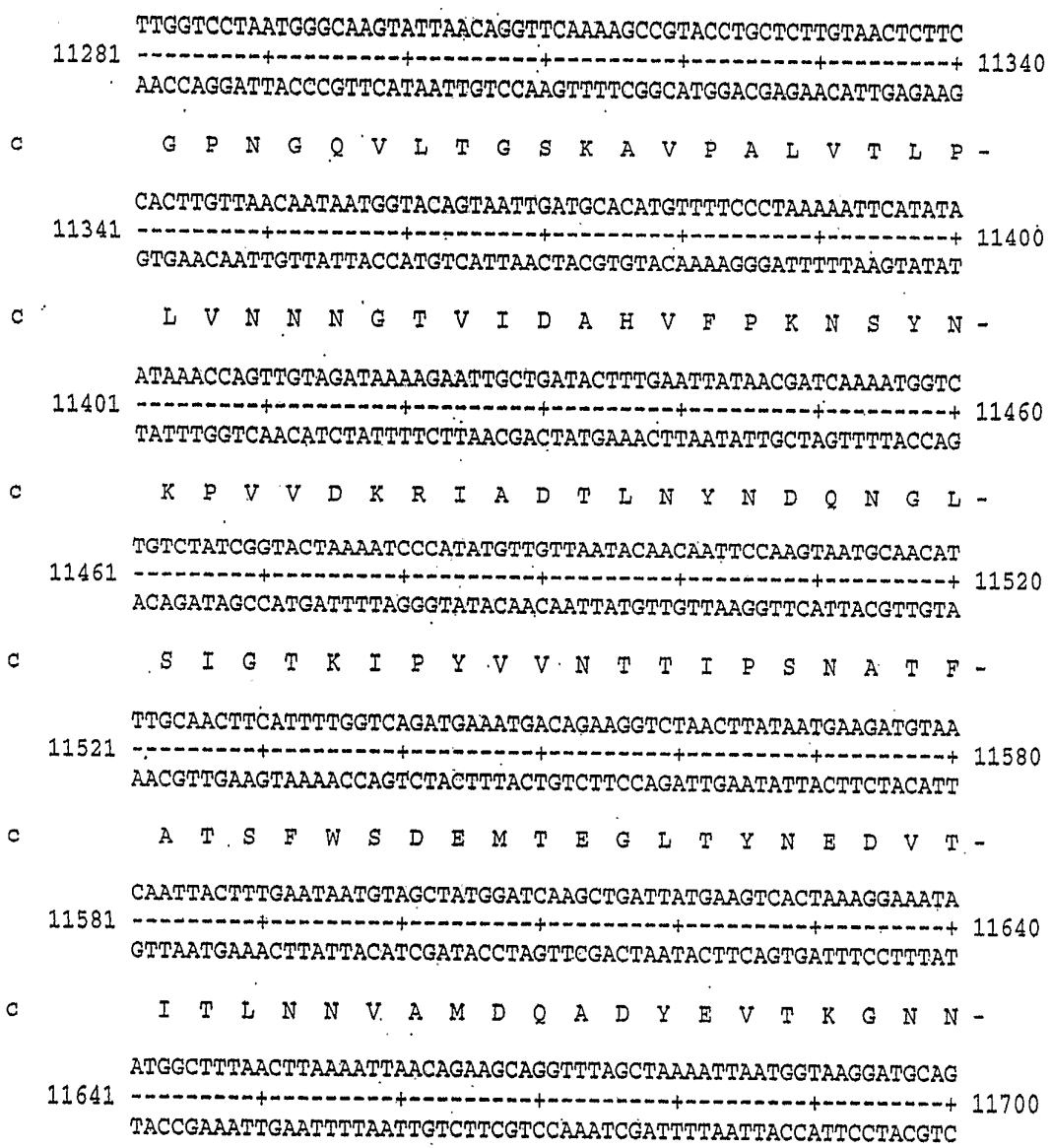


Figure 101AB

```
c          G  F  N  L  K  L  T  E  A  G  L  A  K  I  N  G  K  D  A  D -

       ACCAAAAAATCCAAATTACTTACTCAGCTACTTTGAACTCACTTGCTGTTGCAGACATTC
11701  ---------+---------+---------+---------+---------+---------+ 11760
       TGGTTTTTTAGGTTTAATGAATGAGTCGATGAAACTTGAGTGAACGACAACGTCTGTAAG

c          Q  K  I  Q  I  T  Y  S  A  T  L  N  S  L  A  V  A  D  I  P -

       CTGAAAGTAACGATATTACATATCATTACGGAAATCATCAAGATCATGGGAATACTCCAA
11761  ---------+---------+---------+---------+---------+---------+ 11820
       GACTTTCATTGCTATAATGTATAGTAATGCCTTTAGTAGTTCTAGTACCCTTATGAGGTT

c          E  S  N  D  I  T  Y  H  Y  G  N  H  Q  D  H  G  N  T  P  K -

       AACCAACTAAACCTAATAATGGTCAAATTACAGTAACTAAGACATGGGACAGTCAACCTG
11821  ---------+---------+---------+---------+---------+---------+ 11880
       TTGGTTGATTTGGATTATTACCAGTTTAATGTCATTGATTCTGTACCCTGTCAGTTGGAC

c          P  T  K  P  N  N  G  Q  I  T  V  T  K  T  W  D  S  Q  P  A -

       CTCCTGAGGGGGTGAAAGCGACTGTTCAACTTGTAAATGCCAAGACTGGTGAGAAAGTCG
11881  ---------+---------+---------+---------+---------+---------+ 11940
       GAGGACTCCCCCACTTTCGCTGACAAGTTGAACATTTACGGTTCTGACCACTCTTTCAGC

c          P  E  G  V  K  A  T  V  Q  L  V  N  A  K  T  G  E  K  V  G -

       GTGCTCCTGTAGAACTTTCAGAAAATAATTGGACATATACTTGGAGTGGTCTAGATAATT
11941  ---------+---------+---------+---------+---------+---------+ 12000
       CACGAGGACATCTTGAAAGTCTTTTATTAACCTGTATATGAACCTCACCAGATCTATTAA

c          A  P  V  E  L  S  E  N  N  W  T  Y  T  W  S  G  L  D  N  S -

       CTATTGAATACAAAGTTGAAGAAGAATATAATGGATACTCAGCTGAATACACAGTAGAGA
12001  ---------+---------+---------+---------+---------+---------+ 12060
       GATAACTTATGTTTCAACTTCTTCTTATATTACCTATGAGTCGACTTATGTGTCATCTCT

c          I  E  Y  K  V  E  E  E  Y  N  G  Y  S  A  E  Y  T  V  E  S -

       GCAAAGGGAAGTTGGGGGTAAAAAACTGGAAAGATAATAACCCAGCTCCAATCAATCCTG
12061  ---------+---------+---------+---------+---------+---------+ 12120
```

Figure 101AC

```
          CGTTTCCCTTCAACCCCCATTTTTTGACCTTTCTATTATTGGGTCGAGGTTAGTTAGGAC

c           K  G  K  L  G  V  K  N  W  K  D  N  N  P  A  P  I  N  P  E -

          AAGAACCACGTGTAAAAACATACGGTAAAAAGTTTGTCAAAGTAGACCAAAAAGATACTC
   12121  ---------+---------+---------+---------+---------+---------+  12180
          TTCTTGGTGCACATTTTTGTATGCCATTTTTCAAACAGTTTCATCTGGTTTTTCTATGAG

c           E  P  R  V  K  T  Y  G  K  K  F  V  K  V  D  Q  K  D  T  R -

          GTCTAGAAAATGCGCAGTTCGTTGTTAAAAAAGCAGATAGCAATAAATATATTGCCTTTA
   12181  ---------+---------+---------+---------+---------+---------+  12240
          CAGATCTTTTACGCGTCAAGCAACAATTTTTTCGTCTATCGTTATTTATATAACGGAAAT

c           L  E  N  A  Q  F  V  V  K  K  A  D  S  N  K  Y  I  A  F  K -

          AGTCAACTGCACAACAAGCTGCAGATGAAAAAGCAGCAGCAACTGCAAAACAAAAATTGG
   12241  ---------+---------+---------+---------+---------+---------+  12300
          TCAGTTGACGTGTTGTTCGACGTCTACTTTTTCGTCGTCGTTGACGTTTTGTTTTTAACC

c           S  T  A  Q  Q  A  A  D  E  K  A  A  A  T  A  K  Q  K  L  D -

          ATGCAGCGGTAGCAGCTTACACAAATGCTGCAGATAAGCAAGCCGCTCAAGCTCTAGTAG
   12301  ---------+---------+---------+---------+---------+---------+  12360
          TACGTCGCCATCGTCGAATGTGTTTACGACGTCTATTCGTTCGGCGAGTTCGAGATCATC

c           A  A  V  A  A  Y  T  N  A  A  D  K  Q  A  A  Q  A  L  V  D -

          ATCAAGCACAGCAAGAATACAATGTAGCTTACAAAGAAGCCAAATTTGGTTATGTTGAAG
   12361  ---------+---------+---------+---------+---------+---------+  12420
          TAGTTCGTGTCGTTCTTATGTTACATCGAATGTTTCTTCGGTTTAAACCAATACAACTTC

c           Q  A  Q  Q  E  Y  N  V  A  Y  K  E  A  K  F  G  Y  V  E  V -

          TAGCTGGAAAAGATGAAGCAATGGTTCTTACTTCTAATACGGATGGTCAATTCCAAATTT
   12421  ---------+---------+---------+---------+---------+---------+  12480
          ATCGACCTTTTCTACTTCGTTACCAAGAATGAAGATTATGCCTACCAGTTAAGGTTTAAA

c           A  G  K  D  E  A  M  V  L  T  S  N  T  D  G  Q  F  Q  I  S -
```

Figure 101AD

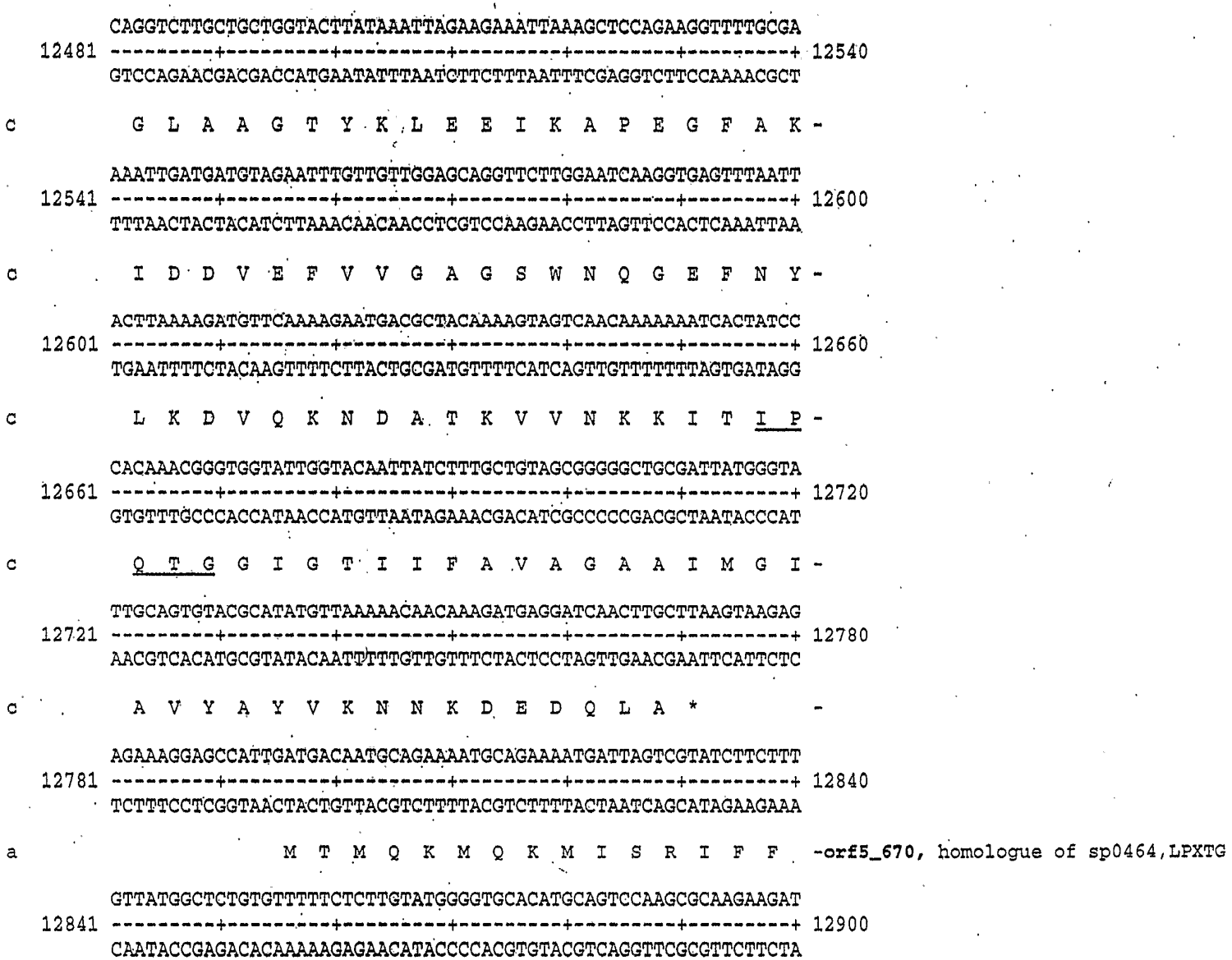


Figure 101AE

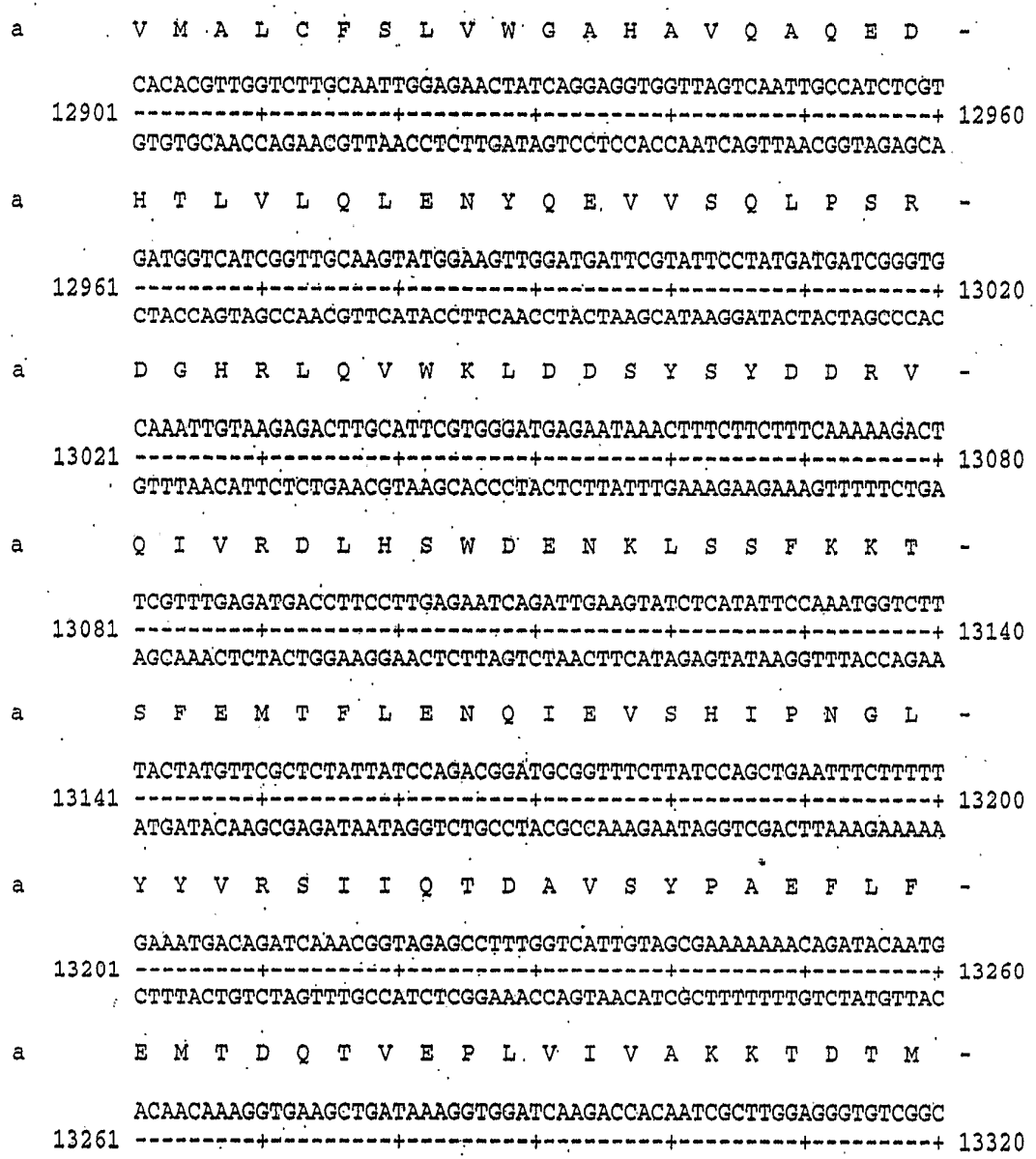


Figure 101AF

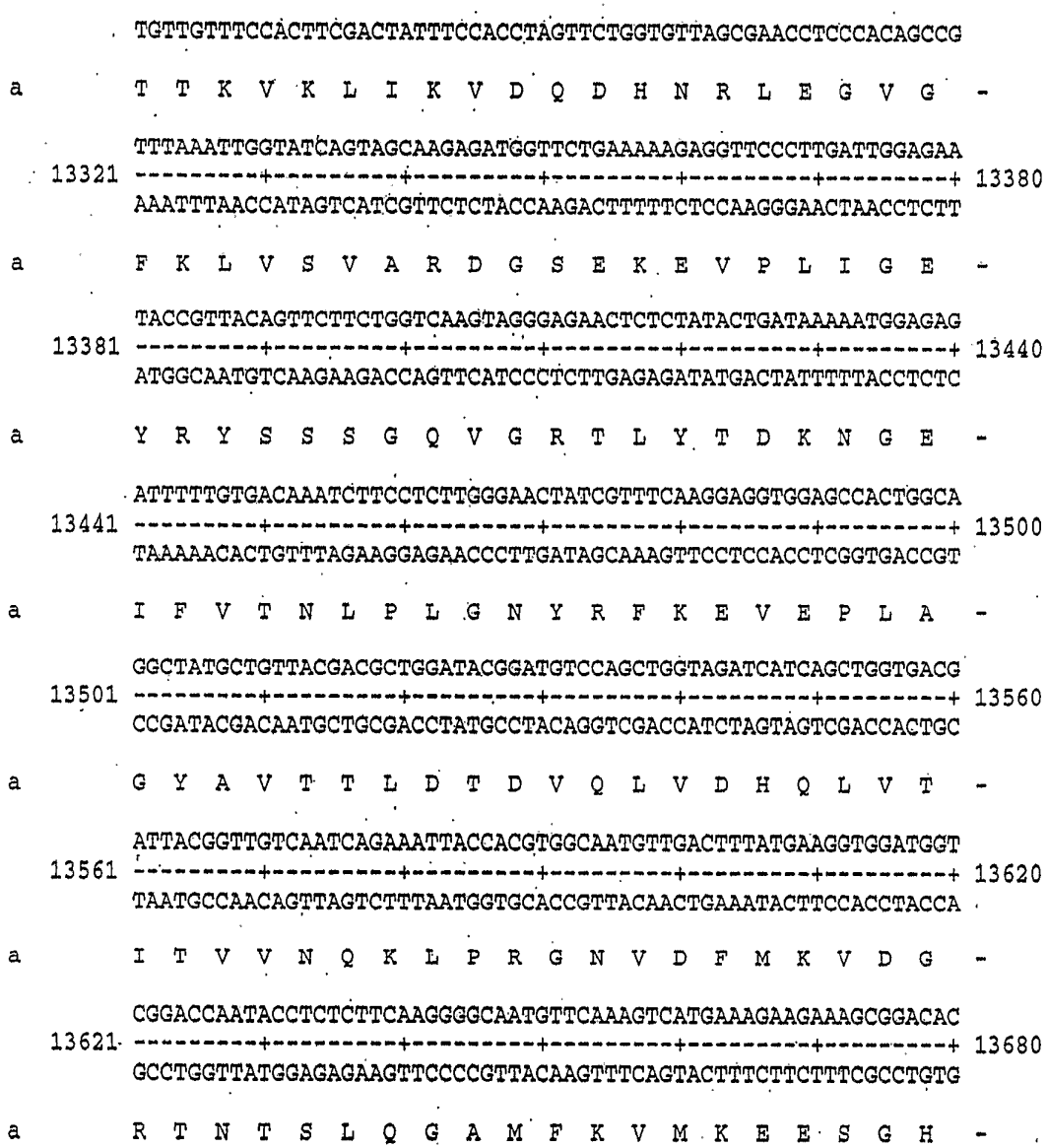


Figure 101AG

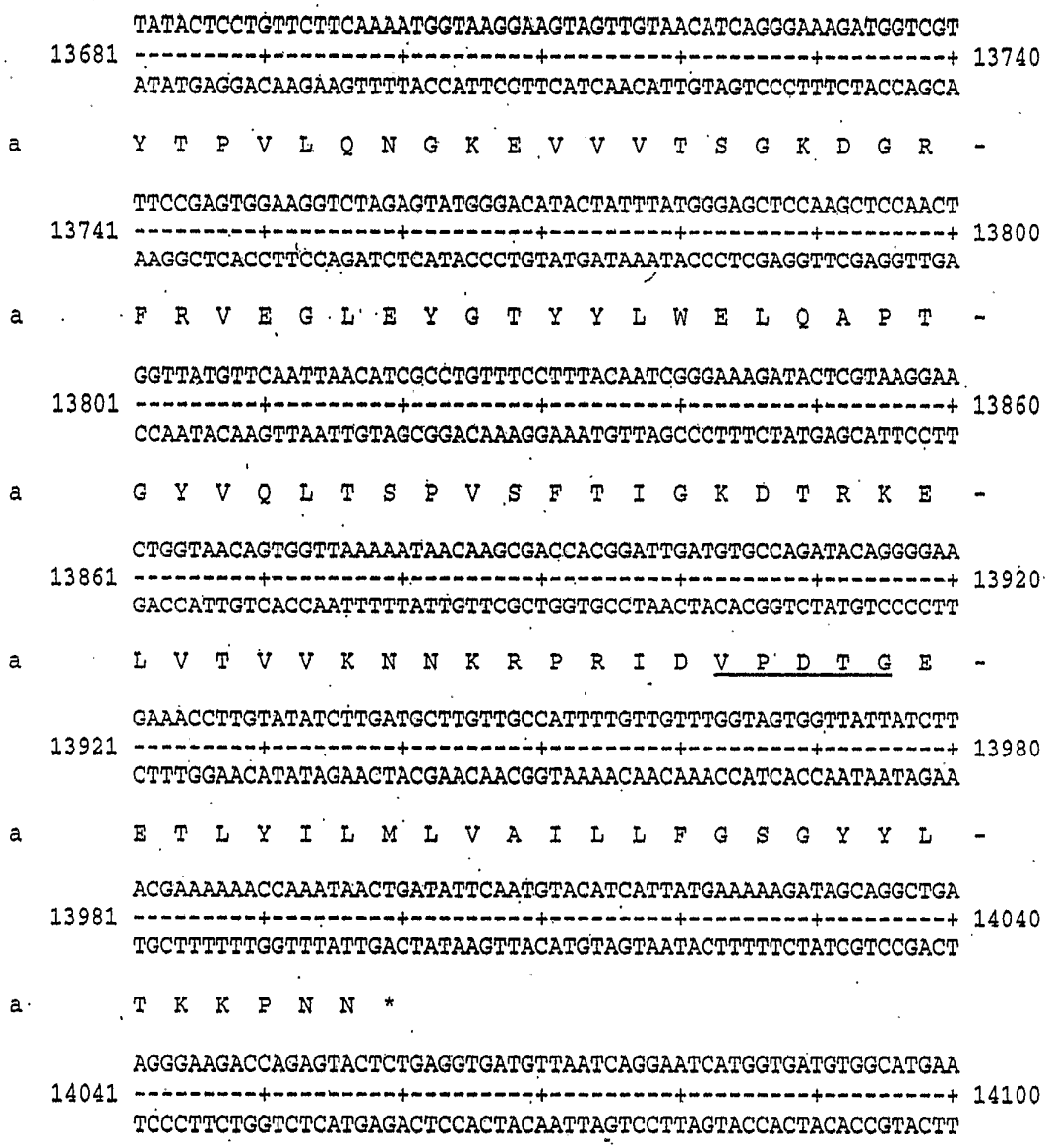


Figure 101AH

```
      TCACAATAACGGATATGAGGCTGGGCAGATTGTGCCAGCCTCATTGTGGGTTATTGTTTG
14101 ---------+---------+---------+---------+---------+---------+ 14160
      AGTGTTATTGCCTATACTCCGACCCGTCTAACACGGTCGGAGTAACACCCAATAACAAAC

      TAAAACGATAGGACTGGTCTGGTAATCATTTTAGGAATGGACAGGACTGGGATTCTGATT
14161 ---------+---------+---------+---------+---------+---------+ 14220
      ATTTTGCTATCCTGACCAGACCATTAGTAAAATCCTTACCTGTCCTGACCCTAAGACTAA

      TAAAATGGATGGTGAATCAGAAAGAAATGAGATTTTCTCGTTTCTCTTAGCAGATAGGAT
14221 ---------+---------+---------+---------+---------+---------+ 14280
      ATTTTACCTACCACTTAGTCTTTCTTTACTCTAAAAGAGCAAAGAGAATCGTCTATCCTA

      TGTCTGTTAGGAAAAGCGATAAAATGATGAGTTTGAAGATAAAGGGATGCTGATAAAAAT
14281 ---------+---------+---------+---------+---------+---------+ 14340
      ACAGACAATCCTTTTCGCTATTTTACTACTCAAACTTCTATTTCCCTACGACTATTTTTA

                                                  M  L  I  K  M  -orf6_670, homologue of sp0466, sortase

      GGTAAAAACAAAAAAGCAAAAACGAAATAATCTCCTATTAGGAGTGGTATTTTTCATTGG
14341 ---------+---------+---------+---------+---------+---------+ 14400
      CCATTTTTGTTTTTTCGTTTTTGCTTTATTAGAGGATAATCCTCACCATAAAAAGTAACC

b      V  K  T  K  K  Q  K  R  N  N  L  L  L  G  V  V  F  F  I  G  -

      AATGGCGGTAATGGCGTATCCGCTGGTGTCTCGCTTGTATTATCGAGTGGAATCAAATCA
14401 ---------+---------+---------+---------+---------+---------+ 14460
      TTACCGCCATTACCGCATAGGCGACCACAGAGCGAACATAATAGCTCACCTTAGTTTAGT

b      M  A  V  M  A  Y  P  L  V  S  R  L  Y  Y  R  V  E  S  N  Q  -

      ACAAATTGCTGACTTTGATAAGGAAAAAGCAACGTTGGATGAGGCTGACATTGATGAACG
14461 ---------+---------+---------+---------+---------+---------+ 14520
      TGTTTAACGACTGAAACTATTCCTTTTTCGTTGCAACCTACTCCGACTGTAACTACTTGC

b      Q  I  A  D  F  D  K  E  K  A  T  L  D  E  A  D  I  D  E  R  -

      AATGAAATTGGCACAAGCCTTCAATGACTCTTTGAATAATGTAGTGAGTGGCGATCCTTG
```

Figure 101AI

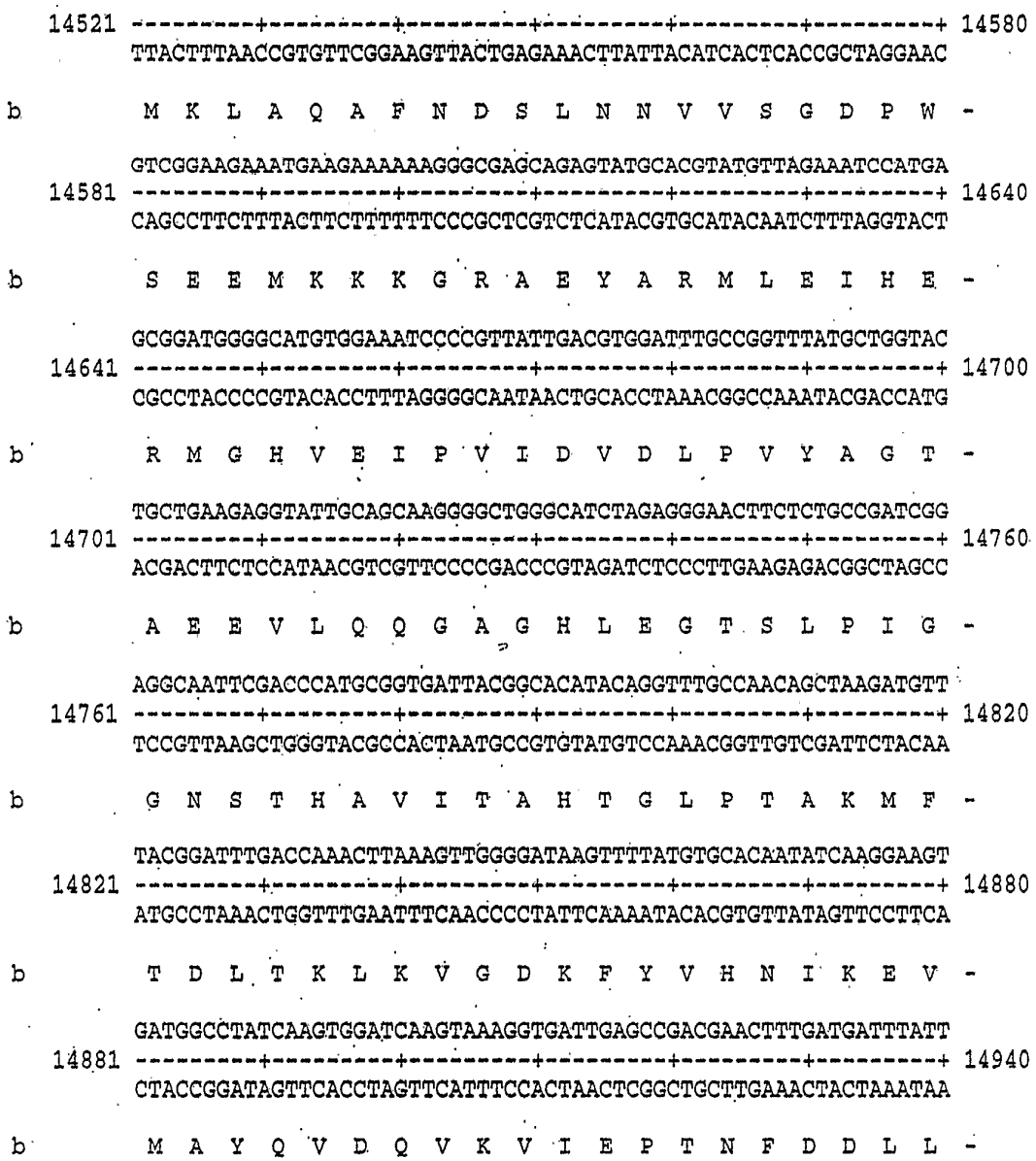


Figure 101AJ

```
          GATTGTACCAGGTCATGATTATGTGACCTTGCTGACTTGTACGCCATACATGATCAATAC
   14941  ---------+---------+---------+---------+---------+---------+  15000
          CTAACATGGTCCAGTACTAATACACTGGAACGACTGAACATGCGGTATGTACTAGTTATG

b          I  V  P  G  H  D  Y  V  T  L  L  T  C  T  P  Y  M  I  N  T  -

          CCATCGTCTATTGGTTCGGGGGCATCGGATACCGTACGTAGCAGAGGTTGAGGAAGAATT
   15001  ---------+---------+---------+---------+---------+---------+  15060
          GGTAGCAGATAACCAAGCCCCCGTAGCCTATGGCATGCATCGTCTCCAACTCCTTCTTAA

b          H  R  L  L  V  R  G  H  R  I  P  Y  V  A  E  V  E  E  E  F  -

          TATTGCAGCAAACAAACTCAGTCATCTCTATCGCTACCTGTTTTATGTGGCAGTTGGTTT
   15061  ---------+---------+---------+---------+---------+---------+  15120
          ATAACGTCGTTTGTTTGAGTCAGTAGAGATAGCGATGGACAAAATACACCGTCAACCAAA

b          I  A  A  N  K  L  S  H  L  Y  R  Y  L  F  Y  V  A  V  G  L  -

          GATTGTGATTCTTTTATGGATTATTCGACGCTTGCGCAAGAAGAAAAAACAACCGGAAAA
   15121  ---------+---------+---------+---------+---------+---------+  15180
          CTAACACTAAGAAAATACCTAATAAGCTGCGAACGCGTTCTTCTTTTTTGTTGGCCTTTT

b          I  V  I  L  L  W  I  I  R  R  L  R  K  K  K  K  Q  P  E  K  -

          GGCTTTGAAGGCGCTGAAAGCAGCAAGGAAGGAAGTGAAGGTGGAGGATGGACAACAGTA
   15181  ---------+---------+---------+---------+---------+---------+  15240
          CCGAAACTTCCGCGACTTTCGTCGTTCCTTCCTTCACTTCCACCTCCTACCTGTTGTCAT

b          A  L  K  A  L  K  A  A  R  K  E  V  K  V  E  D  G  Q  Q  *  -

          GACGTTCACGAAAAAAAGGCACAAAAAAGAAGAAACATCCGCTGATCCTTCTTCTGATTT
   15241  ---------+---------+---------+---------+---------+---------+  15300
          CTGCAAGTGCTTTTTTTCCGTGTTTTTTCTTCTTTGTAGGCGACTAGGAAGAAGACTAAA


          TCTTAGTAGGATTCGCCGTTGCGATATATCCATTGGTGTCTCGTTATTATTATCGTATTG
   15301  ---------+---------+---------+---------+---------+---------+  15360
          AGAATCATCCTAAGCGGCAACGCTATATAGGTAACCACAGAGCAATAATAATAGCATAAC

c                                                  V  S  R  Y  Y  Y  R  I  E  -orf7_670, homologue of sp0467, sortase
```

Figure 101AK

```
      AGTCAAACGAGGTTATTAAAGAGTTTGATGAGACGGTTTCCCAGATGGATAAGGCAGAAC
15361 ---------+---------+---------+---------+---------+---------+ 15420
      TCAGTTTGCTCCAATAATTTCTCAAACTACTCTGCCAAAGGGTCTACCTATTCCGTCTTG

c       S  N  E  V  I  K  E  F  D  E  T  V  S  Q  M  D  K  A  E  L -

      TTGAGGAGCGTTGGCGCTTGGCTCAAGCCTTCAATGCGACCTTGAAACCATCTGAAATTC
15421 ---------+---------+---------+---------+---------+---------+ 15480
      AACTCCTCGCAACCGCGAACCGAGTTCGGAAGTTACGCTGGAACTTTGGTAGACTTTAAG

c       E  E  R  W  R  L  A  Q  A  F  N  A  T  L  K  P  S  E  I  L -

      TTGATCCTTTTACAGAGCAAGAGAAAAAGAAAGGCGTCTCAGAATATGCCAATATGCTAA
15481 ---------+---------+---------+---------+---------+---------+ 15540
      AACTAGGAAAATGTCTCGTTCTCTTTTTCTTTCCGCAGAGTCTTATACGGTTATACGATT

c       D  P  F  T  E  Q  E  K  K  K  G  V  S  E  Y  A  N  M  L  K -

      AGGTCCATGAGCGGATTGGCTATGTGGAAATTCCTGCGATTGATCAGGAAATTCCGATGT
15541 ---------+---------+---------+---------+---------+---------+ 15600
      TCCAGGTACTCGCCTAACCGATACACCTTTAAGGACGCTAACTAGTCCTTTAAGGCTACA

c       V  H  E  R  I  G  Y  V  E  I  P  A  I  D  Q  E  I  P  M  Y -

      ATGTCGGAACGAGTGAGGAAATTCTTCAGAAGGGCGCAGGATTGCTAGAGGGAGCTTCGT
15601 ---------+---------+---------+---------+---------+---------+ 15660
      TACAGCCTTGCTCACTCCTTTAAGAAGTCTTCCCGCGTCCTAACGATCTCCCTCGAAGCA

c       V  G  T  S  E  E  I  L  Q  K  G  A  G  L  L  E  G  A  S  L -

      TACCGGTTGGTGGTGAAAATACCCACACAGTTGTCACTGCTCATAGAGGATTACCGACGG
15661 ---------+---------+---------+---------+---------+---------+ 15720
      ATGGCCAACCACCACTTTTATGGGTGTGTCAACAGTGACGAGTATCTCCTAATGGCTGCC

c       P  V  G  G  E  N  T  H  T  V  V  T  A  H  R  G  L  P  T  A -

      CAGAACTGTTTAGTCAATTGGATAAGATGAAAAAAGGGGATGTCTTTTATCTTCACGTTT
15721 ---------+---------+---------+---------+---------+---------+ 15780
      GTCTTGACAAATCAGTTAACCTATTCTACTTTTTTCCCCTACAGAAAATAGAAGTGCAAA
```

Figure 101AL

```
c          E  L  F  S  Q  L  D  K  M  K  K  G  D  V  F  Y  L  H  V  L -

       TAGACCAGGTGTTGGCCTACCAAGTGGATCAGATTTTGACGGTTGAGCCAAATGACTTTG
15781  ---------+---------+---------+---------+---------+---------+ 15840
       ATCTGGTCCACAACCGGATGGTTCACCTAGTCTAAAACTGCCAACTCGGTTTACTGAAAC

c          D  Q  V  L  A  Y  Q  V  D  Q  I  L  T  V  E  P  N  D  F  E -

       AGCCTGTCTTGATTCAACATGGGGAAGATTATGCGACCTTGTTGACCTGTACACCGTATA
15841  ---------+---------+---------+---------+---------+---------+ 15900
       TCGGACAGAACTAAGTTGTACCCCTTCTAATACGCTGGAACAACTGGACATGTGGCATAT

c          P  V  L  I  Q  H  G  E  D  Y  A  T  L  L  T  C  T  P  Y  M -

       TGATTAACAGTCATCGTCTGTTGGTACGTGGGAAGCGGATTCCGTATACGGCACCAATTG
15901  ---------+---------+---------+---------+---------+---------+ 15960
       ACTAATTGTCAGTAGCAGACAACCATGCACCCTTCGCCTAAGGCATATGCCGTGGTTAAC

c          I  N  S  H  R  L  L  V  R  G  K  R  I  P  Y  T  A  P  I  A -

       CAGAGCGAAATCGAGCGGTGAGAGAGCGTGGGCAATTCTGGTTGTGGTTATTGCTAGCGG
15961  ---------+---------+---------+---------+---------+---------+ 16020
       GTCTCGCTTTAGCTCGCCACTCTCTCGCACCCGTTAAGACCAACACCAATAACGATCGCC

c          E  R  N  R  A  V  R  E  R  G  Q  F  W  L  W  L  L  L  A  A -

       CGTTGGTTATGATTCTGGTATTGAGTTACGGGGTGTATCGTCATCGTCGCATTGTCAAAG
16021  ---------+---------+---------+---------+---------+---------+ 16080
       GCAACCAATACTAAGACCATAACTCAATGCCCCACATAGCAGTAGCAGCGTAACAGTTTC

c          L  V  M  I  L  V  L  S  Y  G  V  Y  R  H  R  R  I  V  K  G -

       GGCTAGAAAAACAATTGGAGGAGCATCATGTCAAAGGCTAAGCTACAGAAATTACTAGGG
16081  ---------+---------+---------+---------+---------+---------+ 16140
       CCGATCTTTTTGTTAACCTCCTCGTAGTACAGTTTCCGATTCGATGTCTTTAATGATCCC

a                                 M  S  K  A  K  L  Q  K  L  L  G   -orf8_670, homologue of sp0468, sortase

c          L  E  K  Q  L  E  E  H  H  V  K  G  *
```

Figure 101AM

```
          TATTTGCTGATGCTGGTAGCATTGGTGATTCCTGTTTATTGTTTTGGGCAGATGGTGTTA
   16141  ---------+---------+---------+---------+---------+---------+  16200
          ATAAACGACTACGACCATCGTAACCACTAAGGACAAATAACAAAACCCGTCTACCACAAT

a         Y  L  L  M  L  V  A  L  V  I  P  V  Y  C  F  G  Q  M  V  L  -

          CAGTCTTTAGGACAAGTAAAAGGTCATGAGATATTTTCAGAATCTGTGACGGCCGACAGT
   16201  ---------+---------+---------+---------+---------+---------+  16260
          GTCAGAAATCCTGTTCATTTTCCAGTACTCTATAAAAGTCTTAGACACTGCCGGCTGTCA

a         Q  S  L  G  Q  V  K  G  H  E  I  F  S  E  S  V  T  A  D  S  -

          TACCAAGAGCAATTGCAACGGTCGCTTGATTACAATCAACGCTTGGATTCGCAAAATCGT
   16261  ---------+---------+---------+---------+---------+---------+  16320
          ATGGTTCTCGTTAACGTTGCCAGCGAACTAATGTTAGTTGCGAACCTAAGCGTTTTAGCA

a         Y  Q  E  Q  L  Q  R  S  L  D  Y  N  Q  R  L  D  S  Q  N  R  -

          ATTGTAGATCCTTTTTTGGCGGAAGGGTATGAGGTAAATTACCAAGTGTCTGACGATCCT
   16321  ---------+---------+---------+---------+---------+---------+  16380
          TAACATCTAGGAAAAAACCGCCTTCCCATACTCCATTTAATGGTTCACAGACTGCTAGGA

a         I  V  D  P  F  L  A  E  G  Y  E  V  N  Y  Q  V  S  D  D  P  -

          GATGCAGTCTACGGCTATTTGTCGATTCCGAGTTTGGAAATCATGGAGCCAGTTTATCTA
   16381  ---------+---------+---------+---------+---------+---------+  16440
          CTACGTCAGATGCCGATAAACAGCTAAGGCTCAAACCTTTAGTACCTCGGTCAAATAGAT

a         D  A  V  Y  G  Y  L  S  I  P  S  L  E  I  M  E  P  V  Y  L  -

          GGAGCGGATTACCATCATTTAGCAATGGGGTTGGCCCATGTGGATGGGACGCCTCTTCCT
   16441  ---------+---------+---------+---------+---------+---------+  16500
          CCTCGCCTAATGGTAGTAAATCGTTACCCCAACCGGGTACACCTACCCTGCGGAGAAGGA

a         G  A  D  Y  H  H  L  A  M  G  L  A  H  V  D  G  T  P  L  P  -

          GTTGAGGGAAAAGGGATTCGTTCAGTGATTGCTGGGCACCGTGCAGAACCAAGCCATGTC
   16501  ---------+---------+---------+---------+---------+---------+  16560
          CAACTCCCTTTTCCCTAAGCAAGTCACTAACGACCCGTGGCACGTCTTGGTTCGGTACAG
```

Figure 101AM

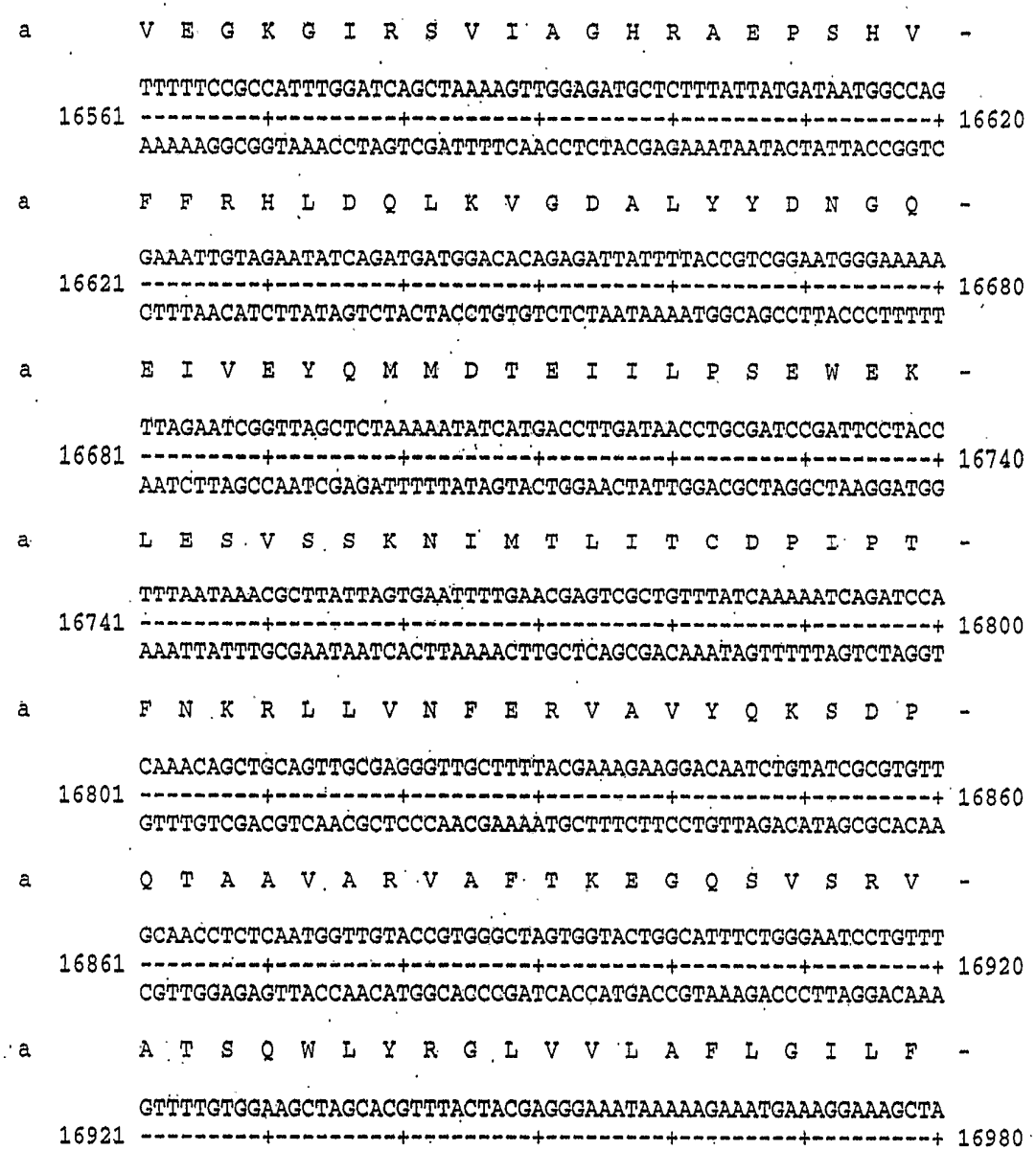


Figure 101AN

```
      CAAAACACCTTCGATCGTGCAAATGATGCTCCCTTTATTTTTCTTTACTTTCCTTTCGAT

a     V  L  W  K  L  A  R  L  L  R  G  K  *

      AGGCTGTTCCTTTTTCCGGCTCTTTGTCAACTGTAGTGGGTTGAAAAAAAGCTAAGCTCG
16981 ---------+---------+---------+---------+---------+---------+ 17040
      TCCGACAAGGAAAAAGGCCGAGAAACAGTTGACATCACCCAACTTTTTTTCGATTCGAGC


      AGAAAGGACAAATTTTGTCCTTTCTTTTTTGATATTCAGAGCGATAAAAATCCGTTTTTT
17041 ---------+---------+---------+---------+---------+---------+ 17100
      TCTTTCCTGTTTAAAACAGGAAAGAAAAAACTATAAGTCTCGCTATTTTTAGGCAAAAAA


      GAAGTTTTCAAA
17101 ----------+-- 17112
      CTTCAAAAGTTT
```

PCT/US05/27239

293/487

M1, strain 2580

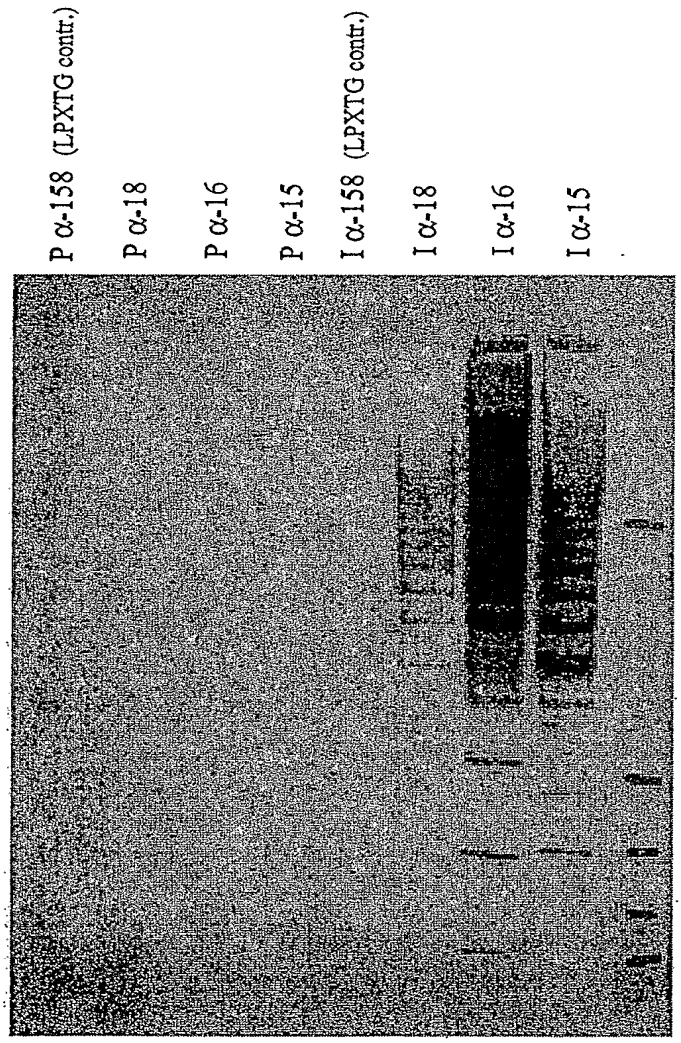


Figure 102

LEGEND:

I α-#: immune serum anti-#
P α-#: pre-immune serum anti-#

M1, strain 2913

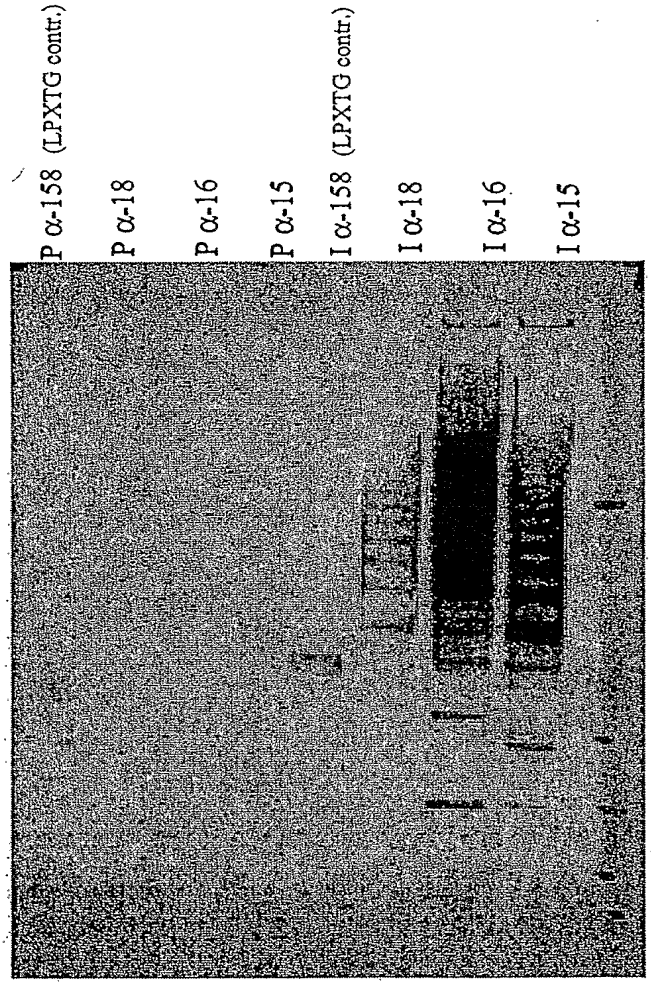
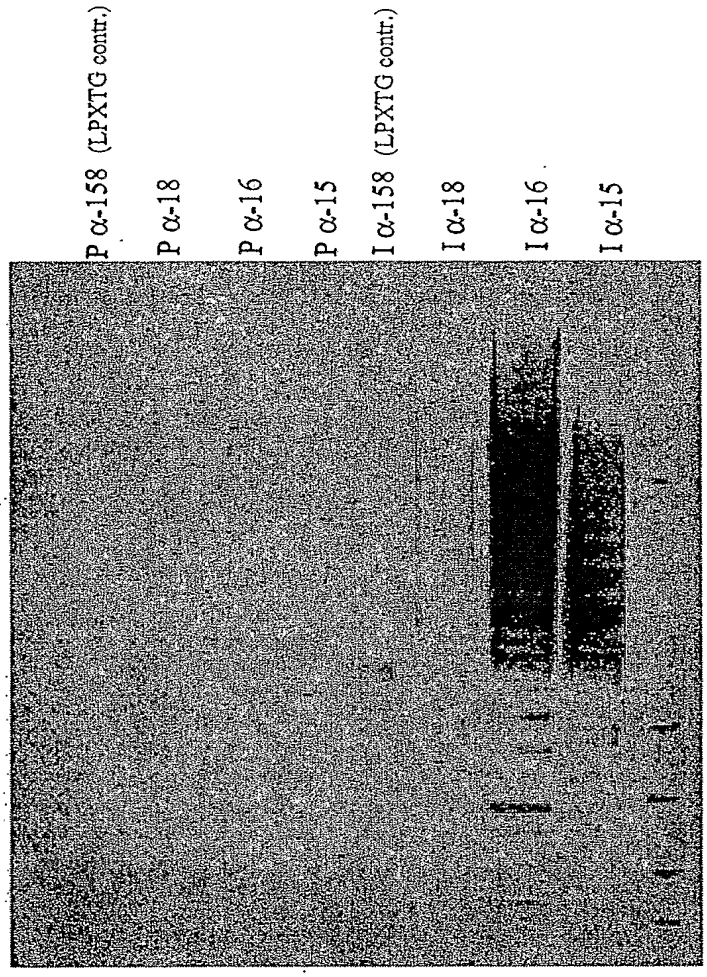


Figure 103

LEGEND:

I α-#: immune serum anti-#
P α-#: pre-immune serum anti-#

M1, strain 3280



LEGEND:

I α-#: immune serum anti-#
P α-#: pre-immune serum anti-#

Figure 104

M1 strain 3348

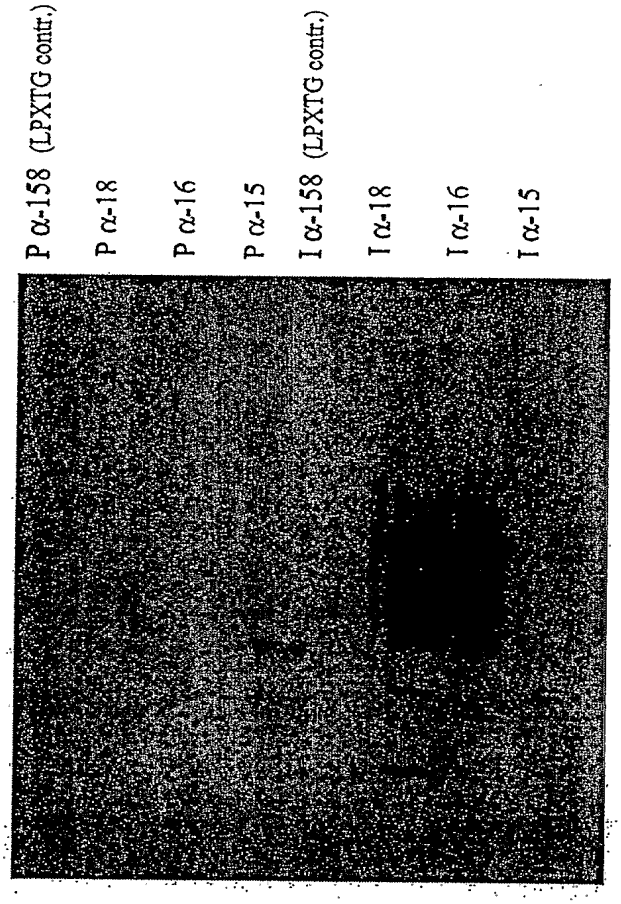


Figure 105

M1 strain 2719

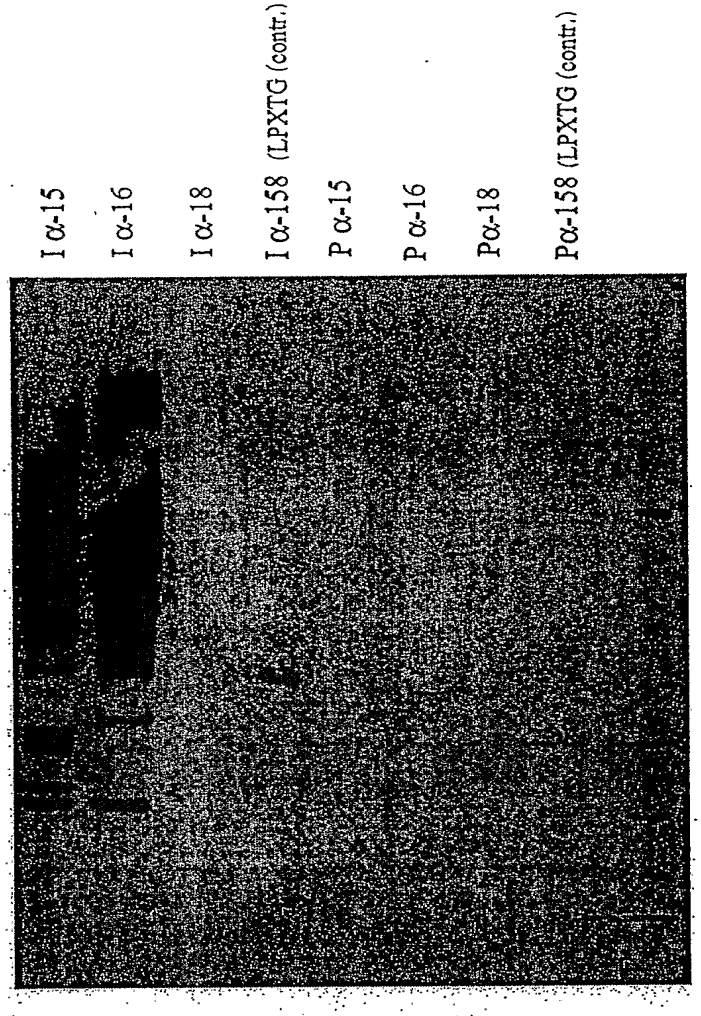


Figure 106

Western blot Western blot on fraction enriched in surface proteins of M1 (SF370)

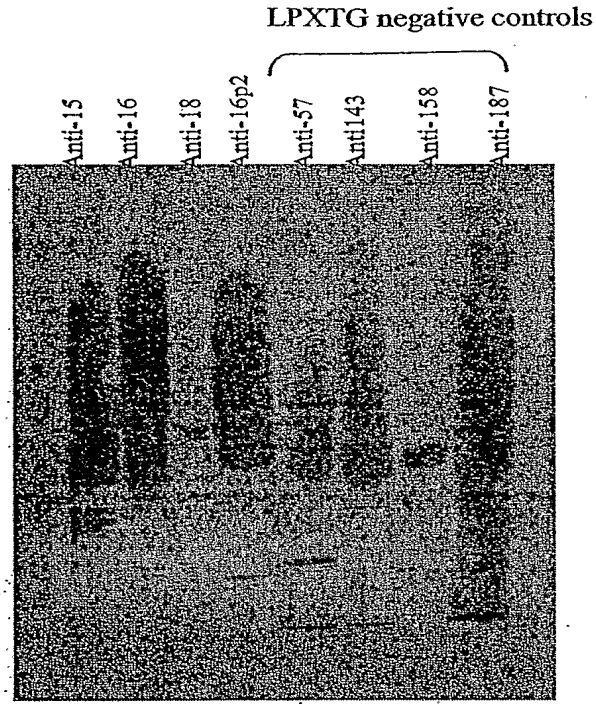
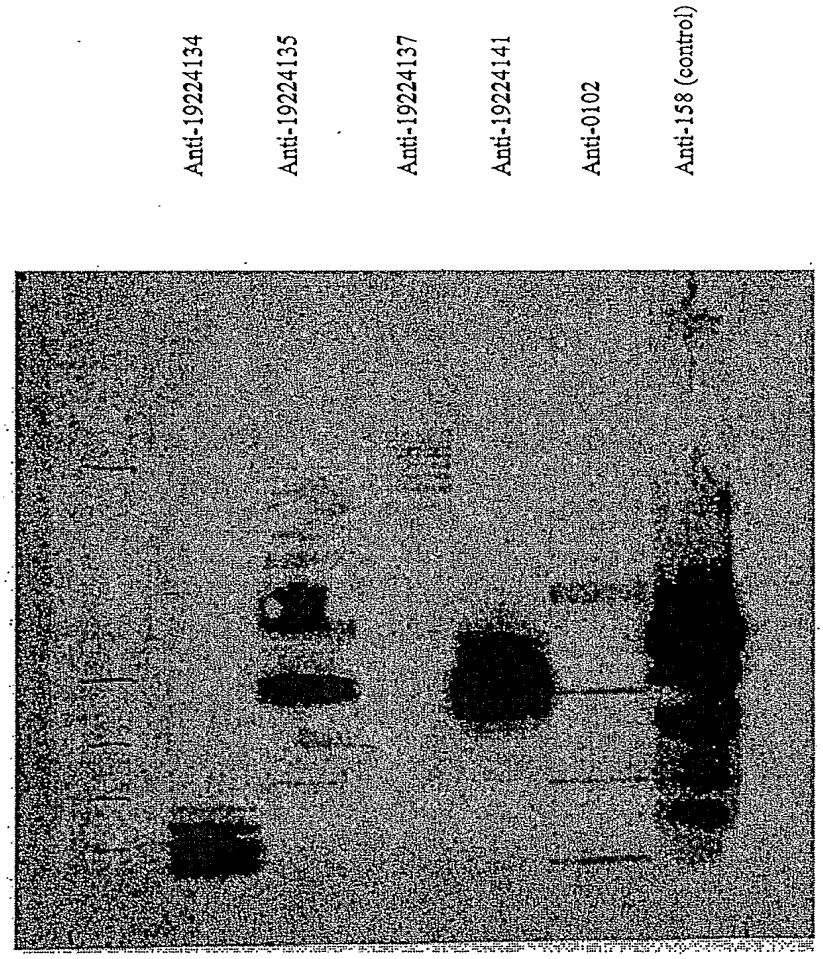


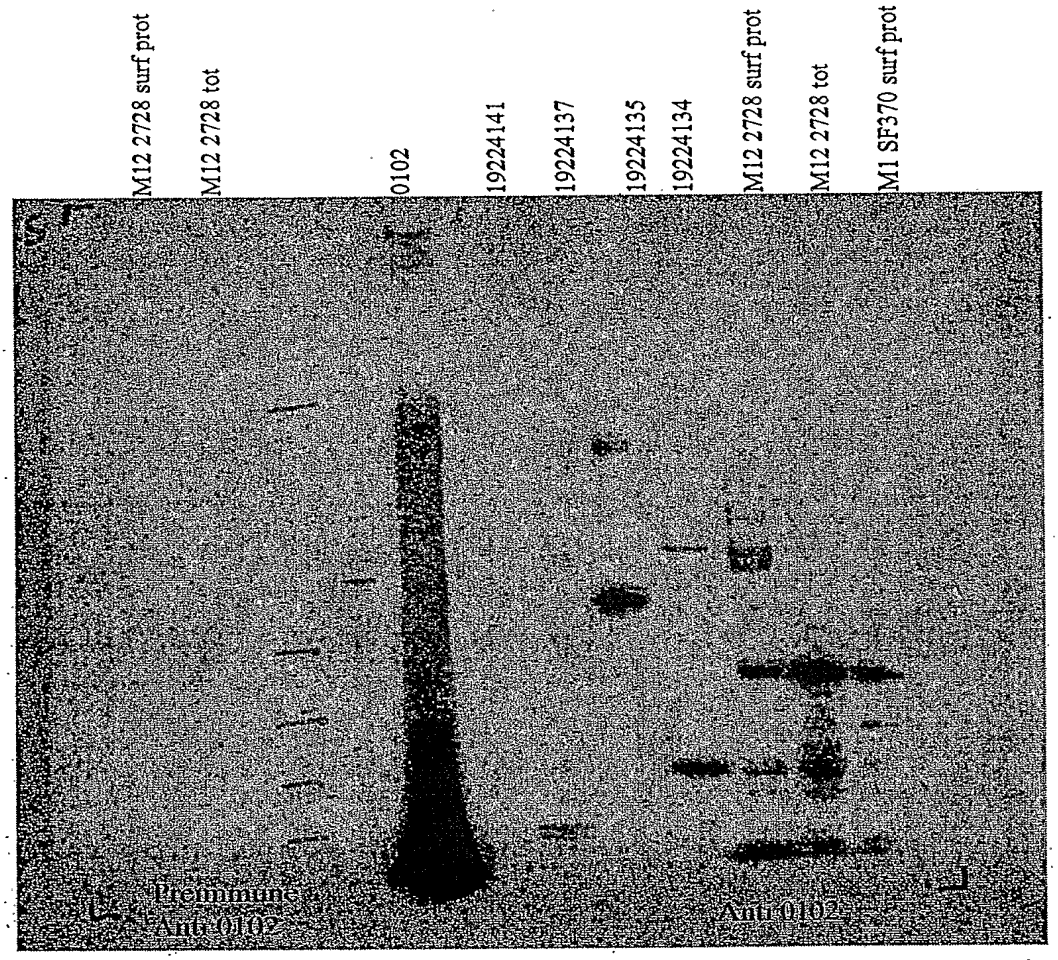
Figure 107

Western blot on fraction enriched in surface proteins of M12 (2728)



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel"WB M12 (23-11-4b2a)"-, the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in M12 (strain isolate 2728). This means that in M12 three out of the five LPKTG proteins form high-molecular weight structures:

Figure 108

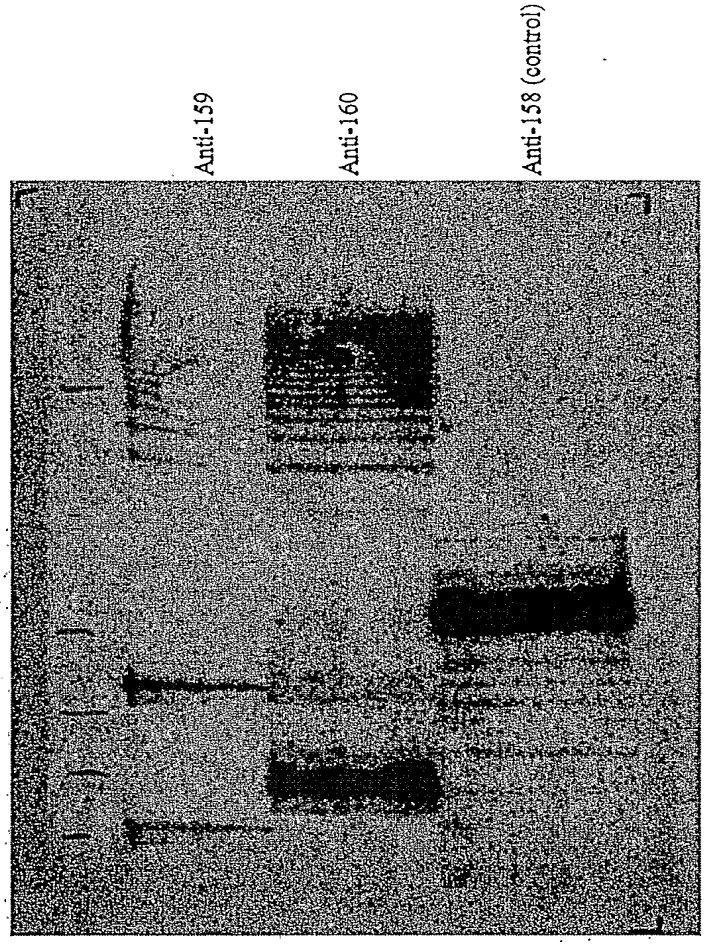


Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel"WB M12 (23-11-4b2bis)"-, the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in protein extracts from M12 (strain isolate 2728.)

Figure 109

PCT/US05/27239

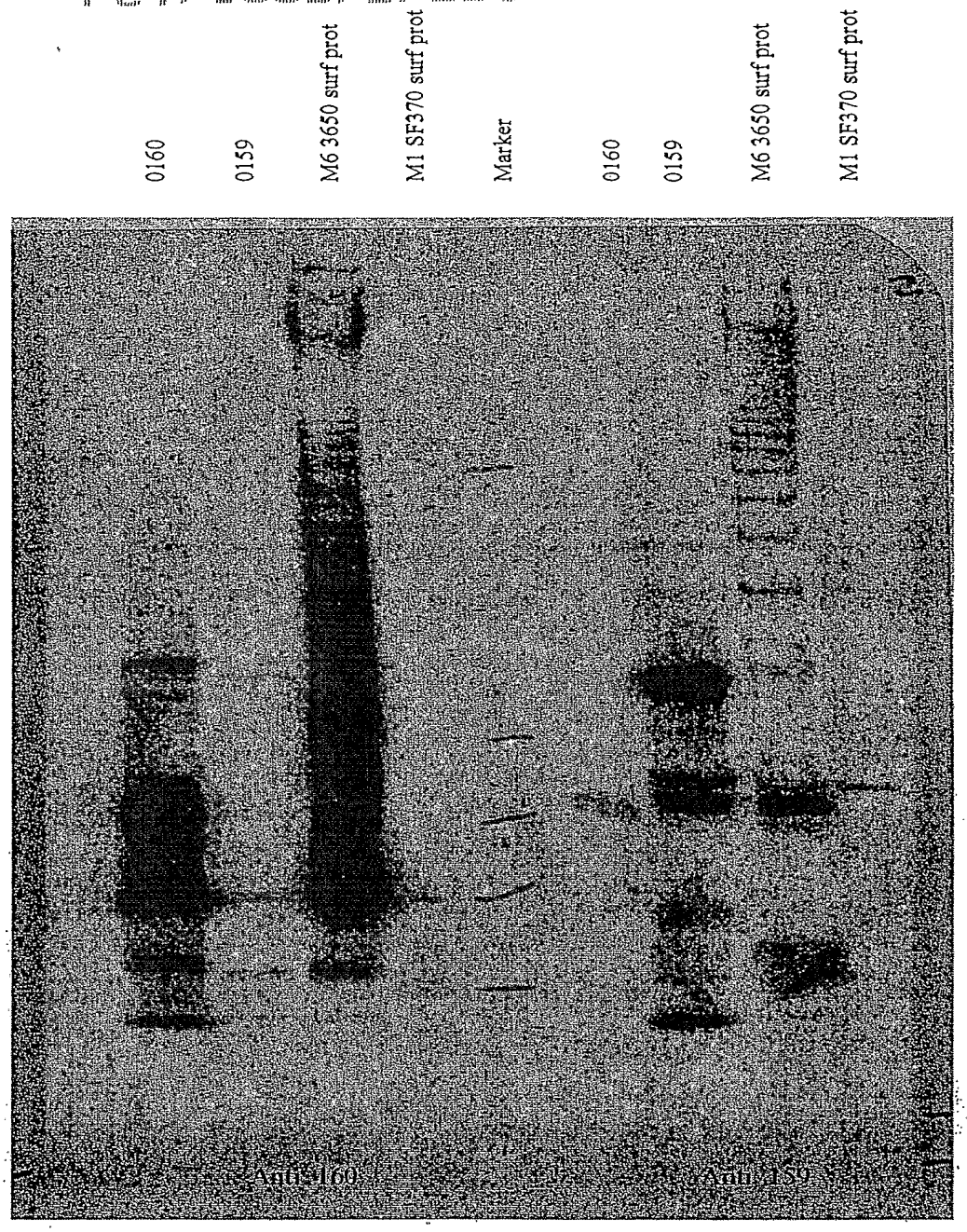
Western blot on fraction enriched in surface proteins of M6 (2724)



M6 strain isolate 2724

Figure 110

PCT/US05/27239



M6 strain isolate 3650

Figure 111

PCT/US05/27239

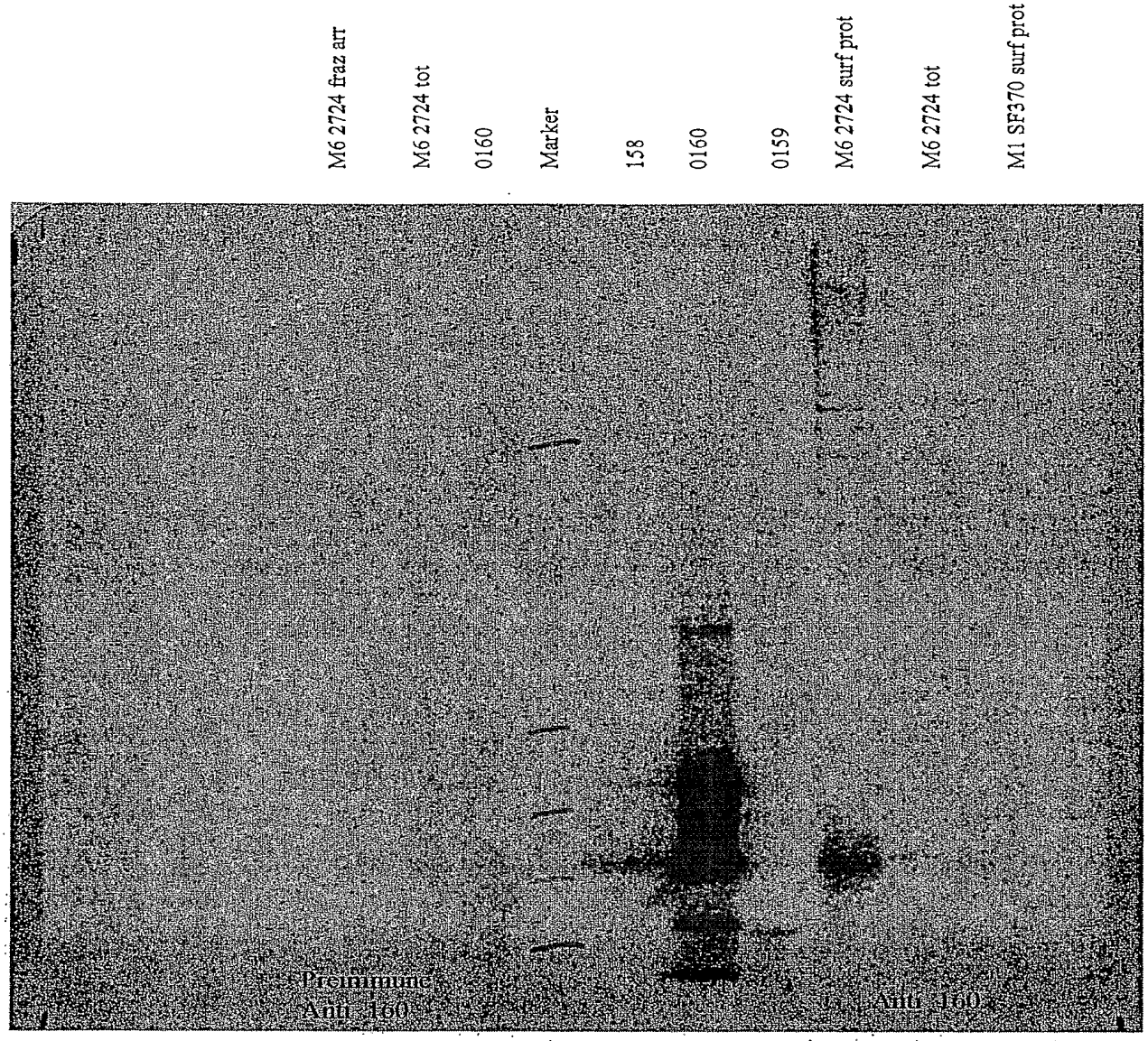


Figure 112

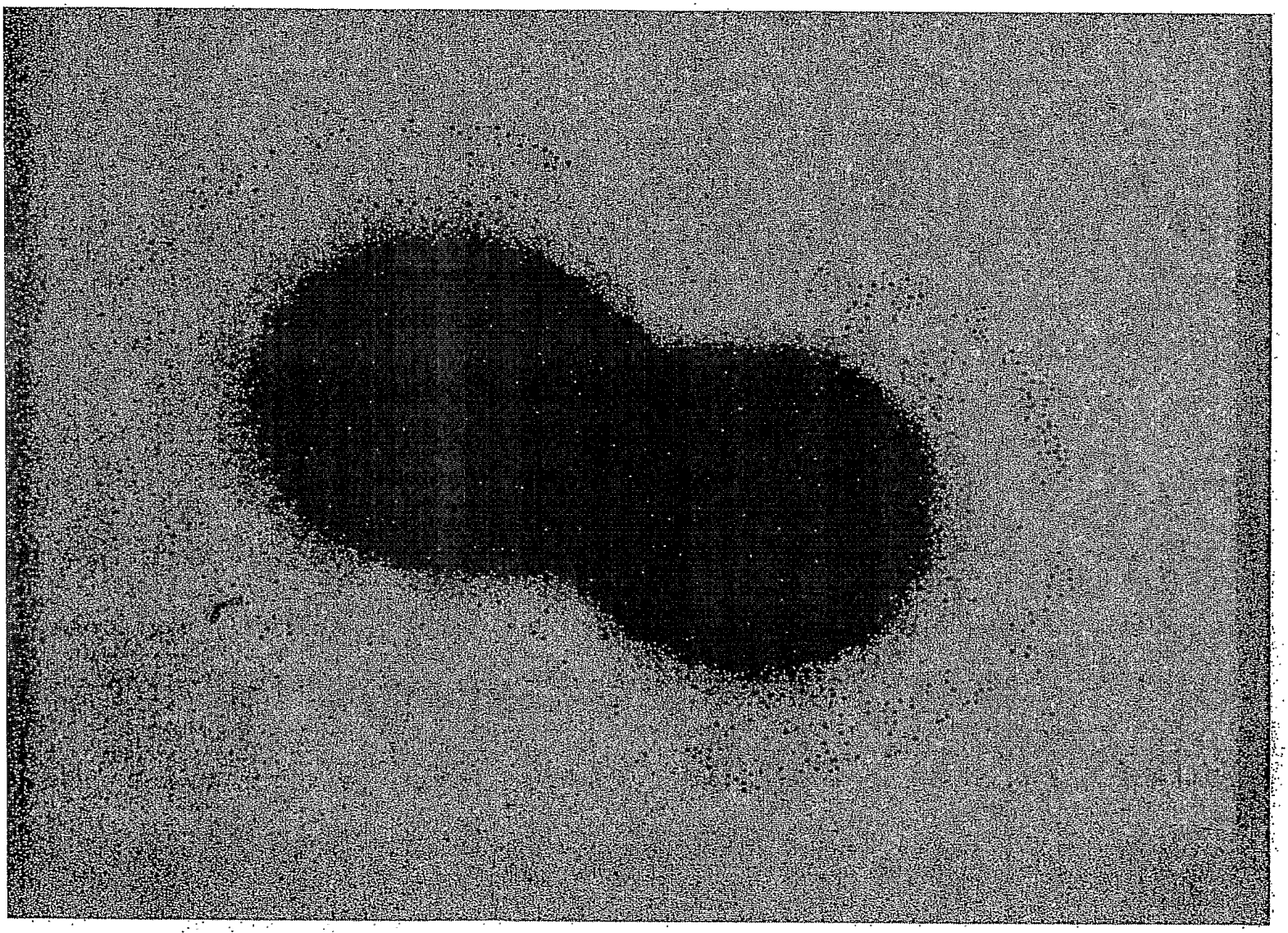


FIGURE 113

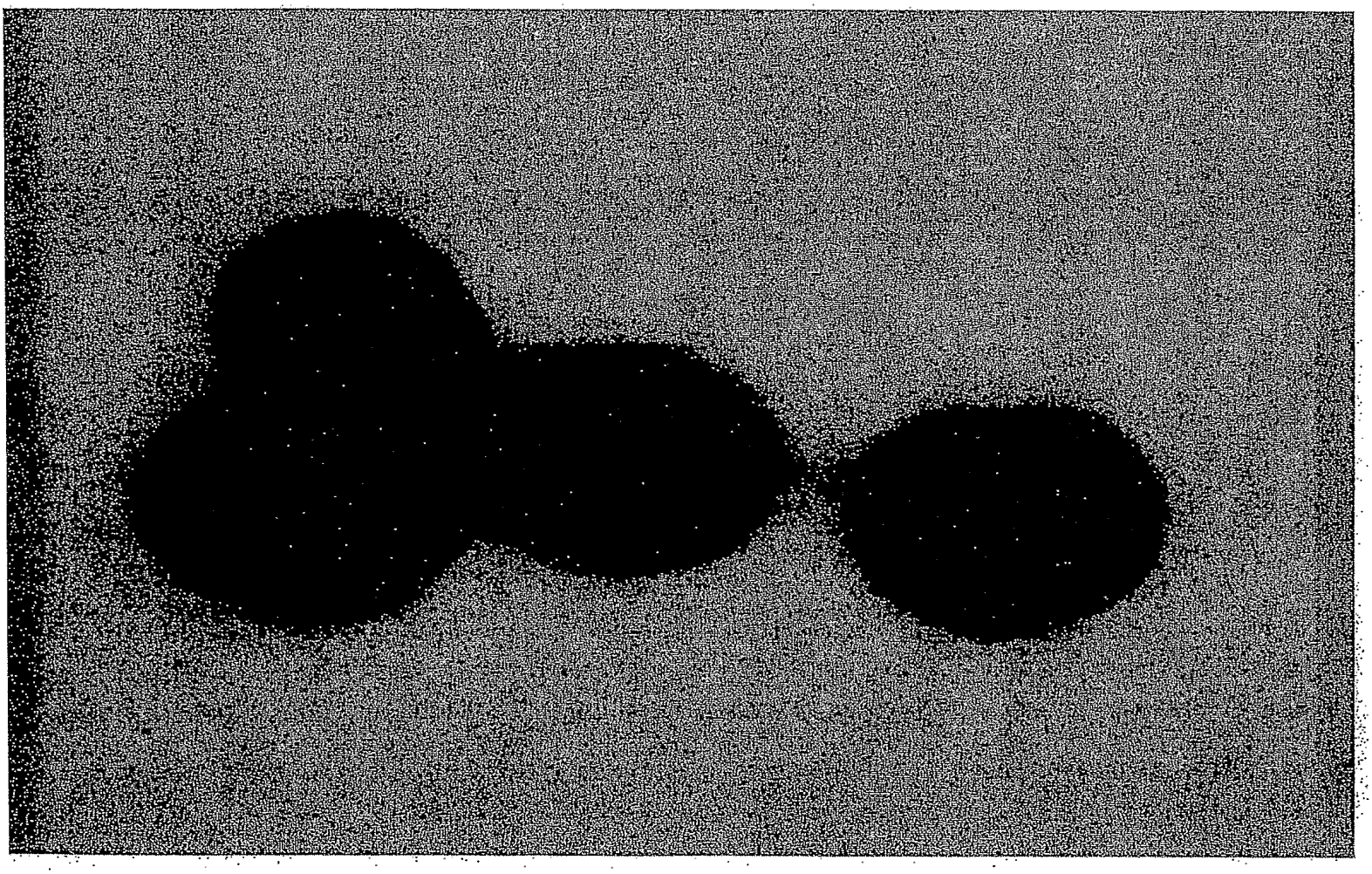


FIGURE 114

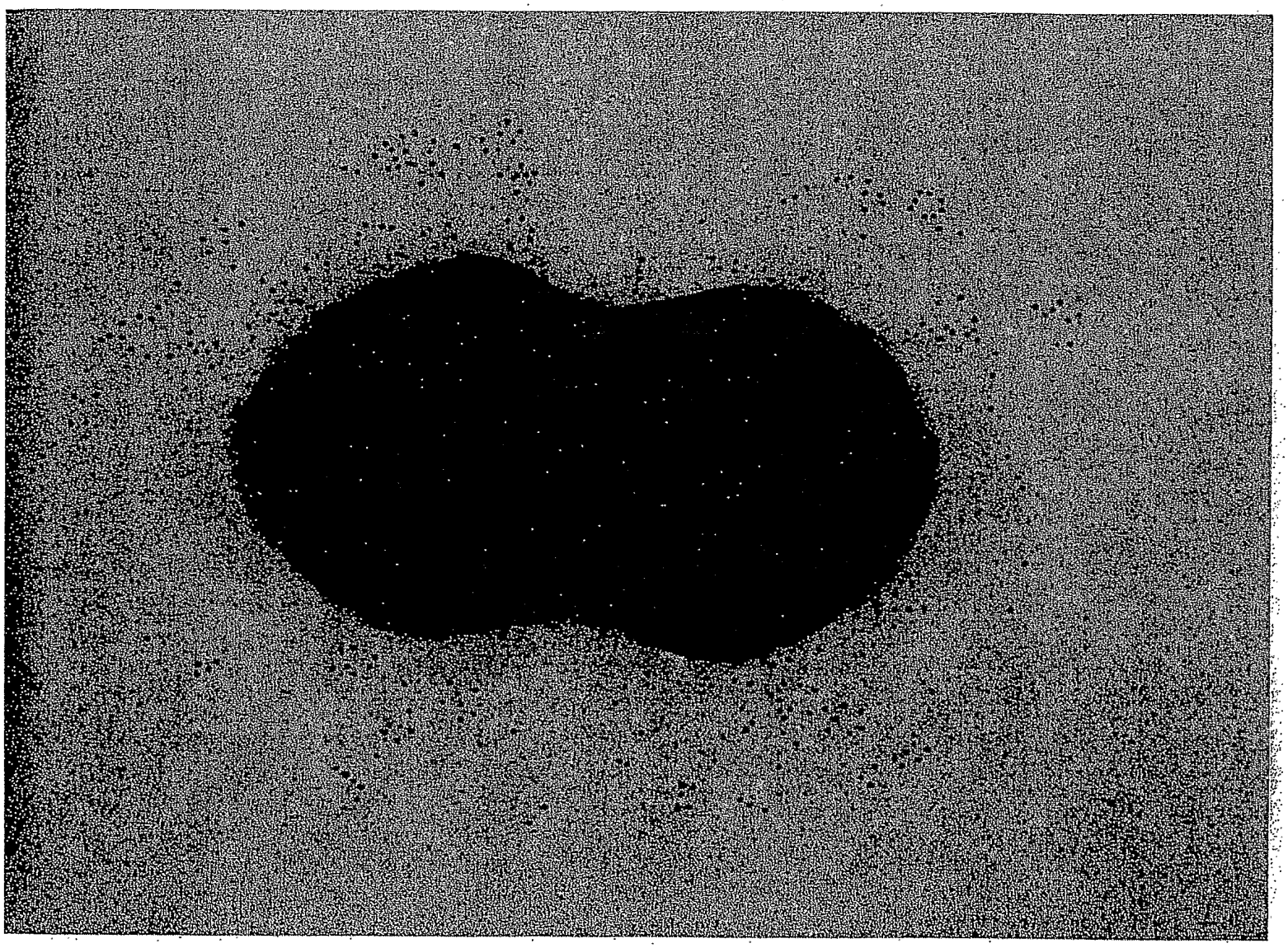


FIGURE 115

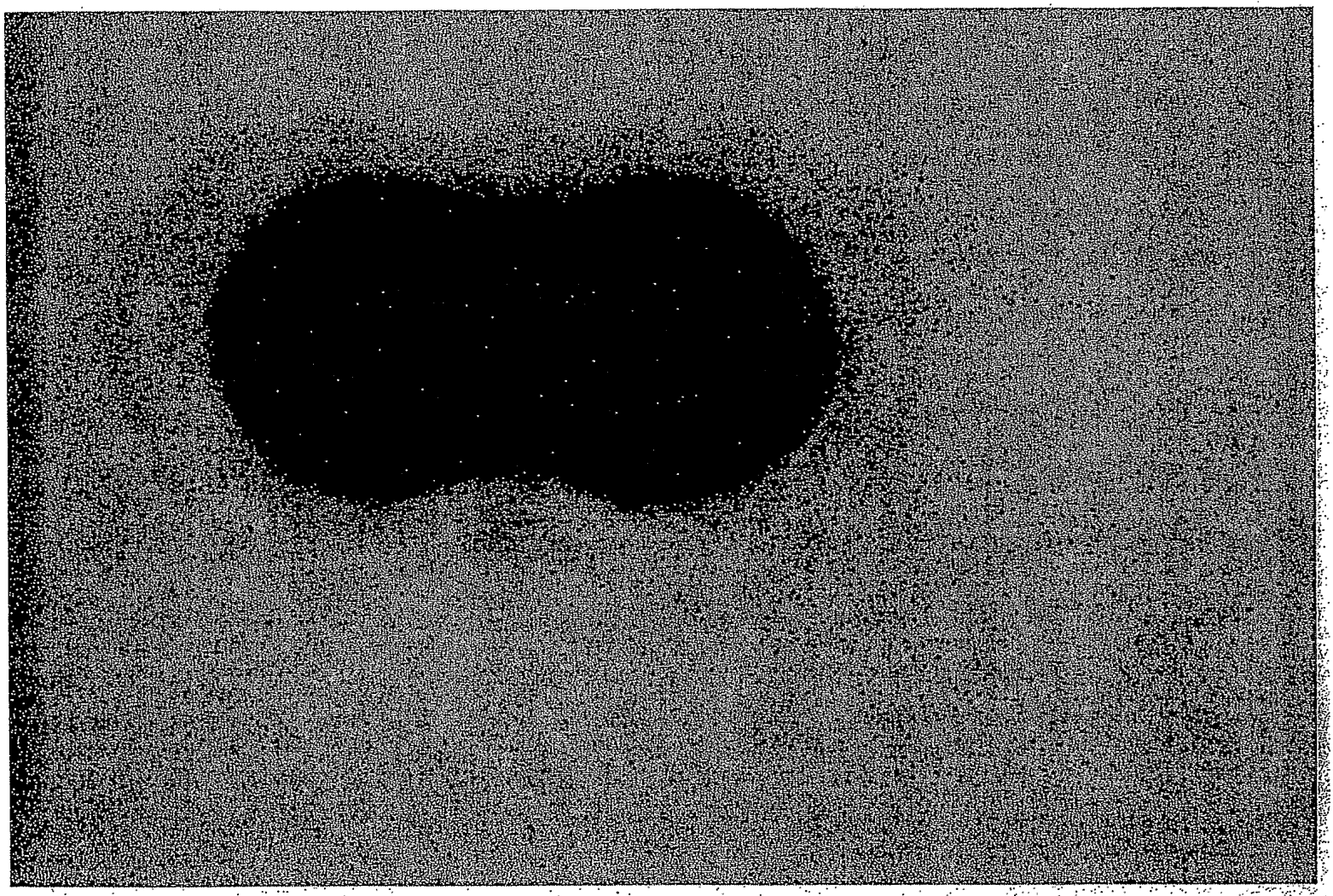


FIGURE 116

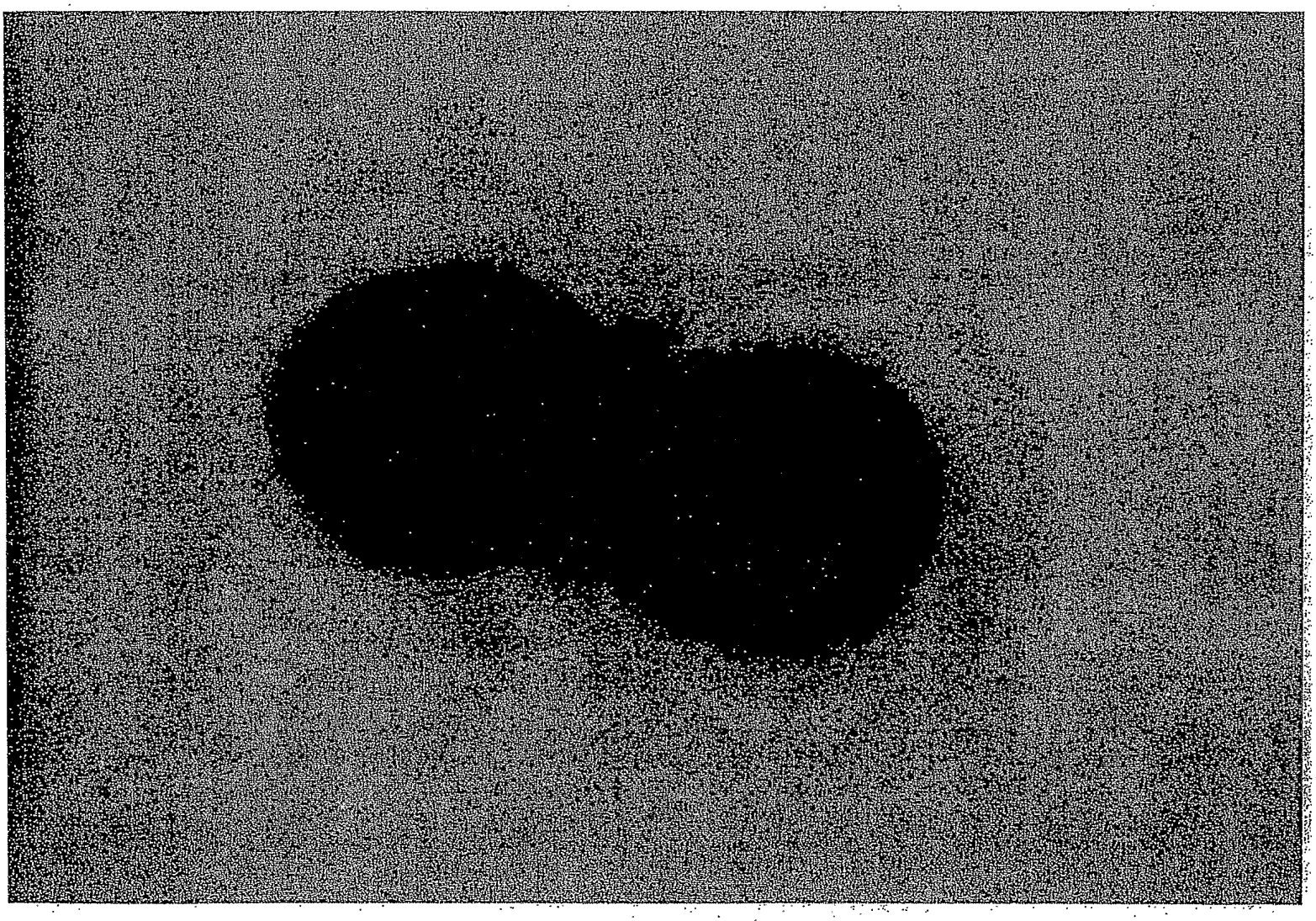


FIGURE 117

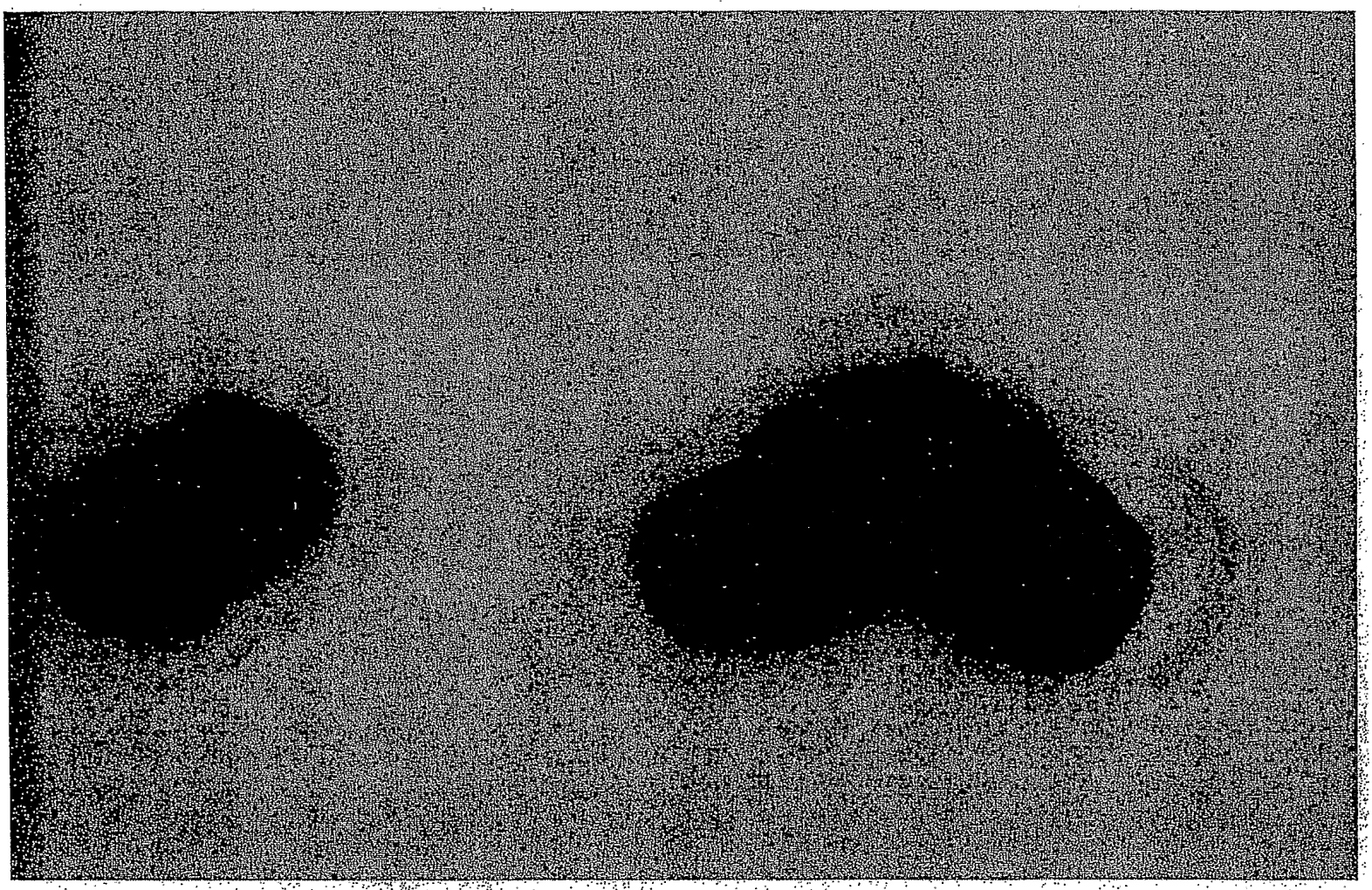


FIGURE 118

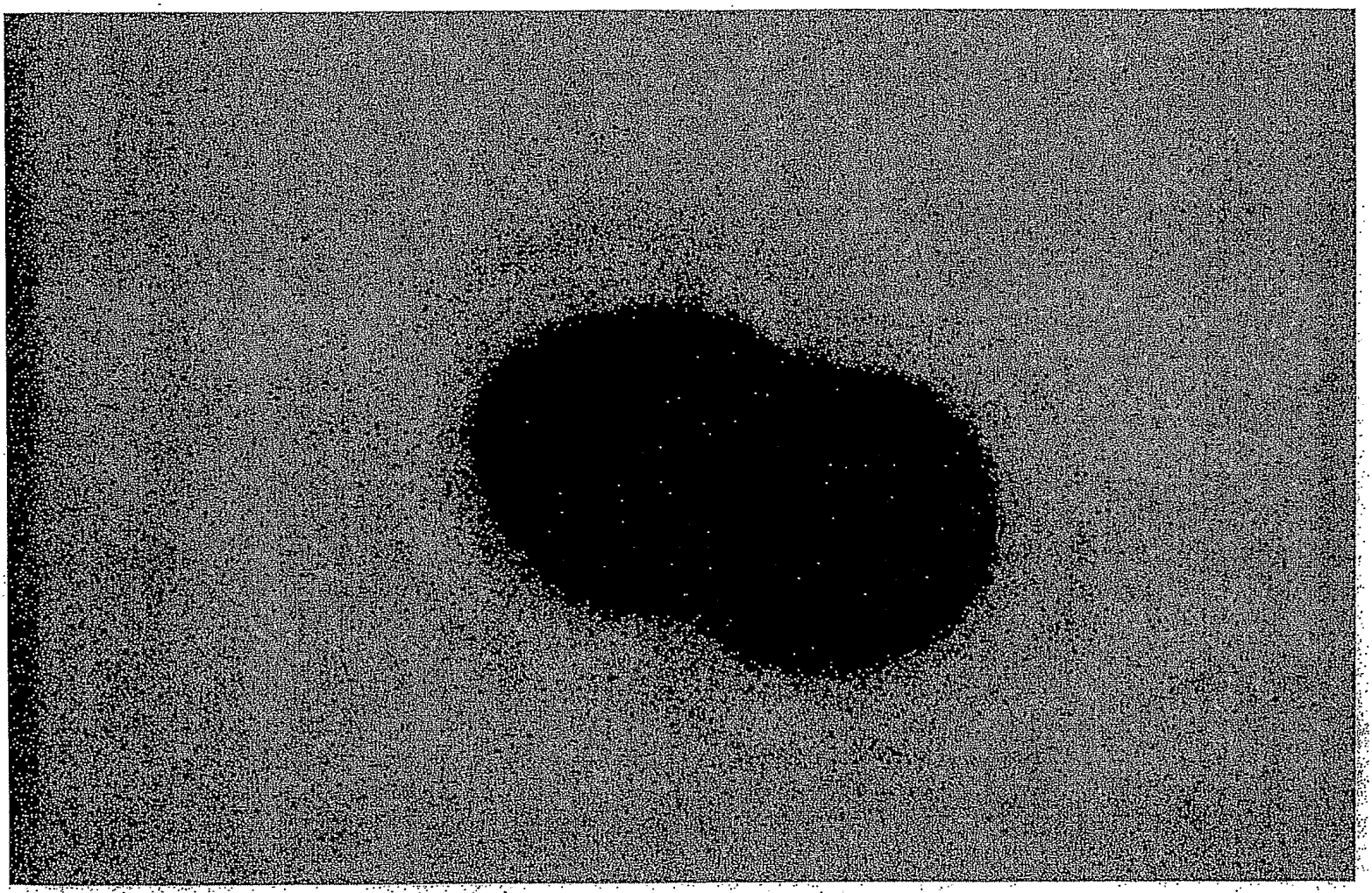


FIGURE 119

PCT/US05/27239

311/487

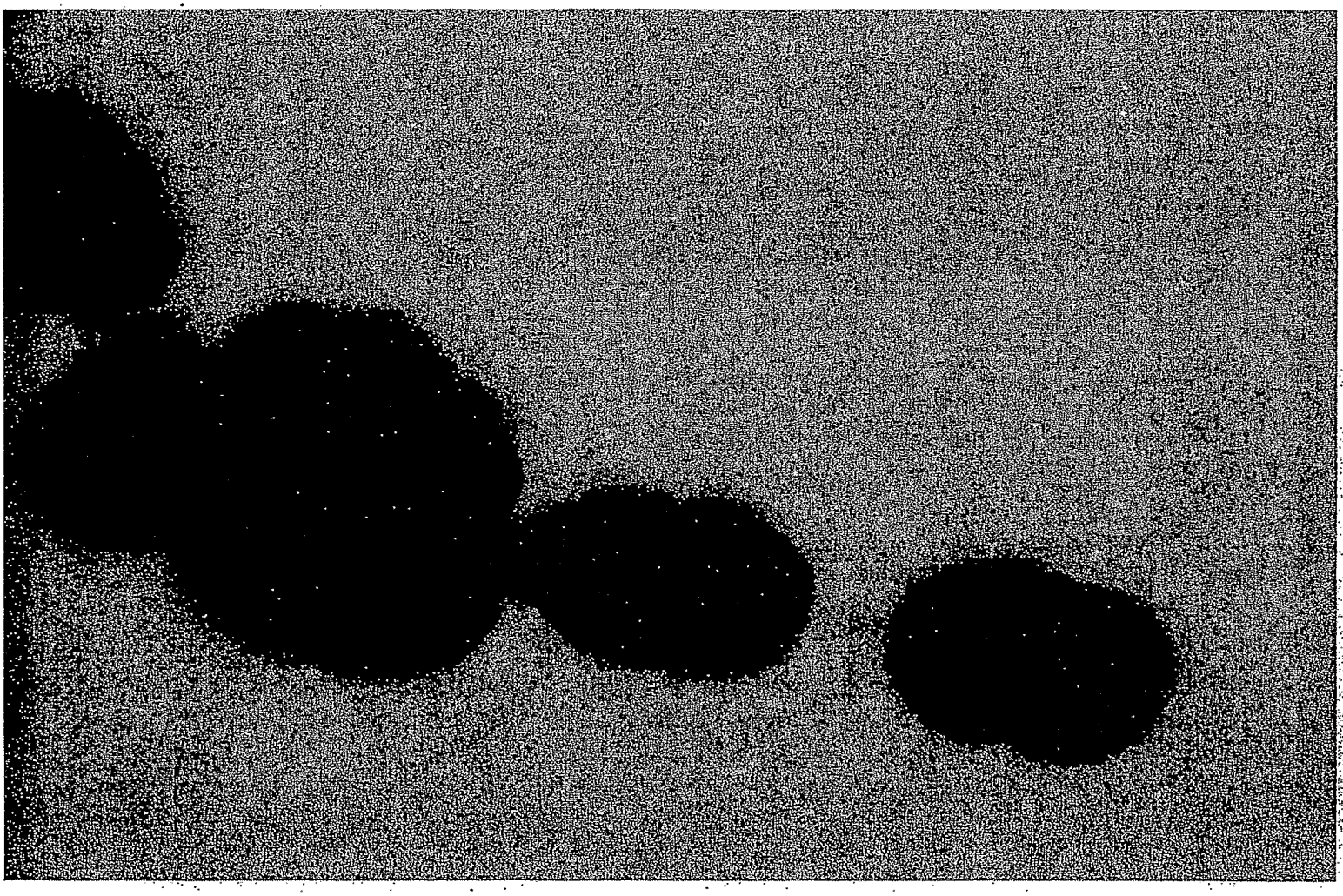


FIGURE 120

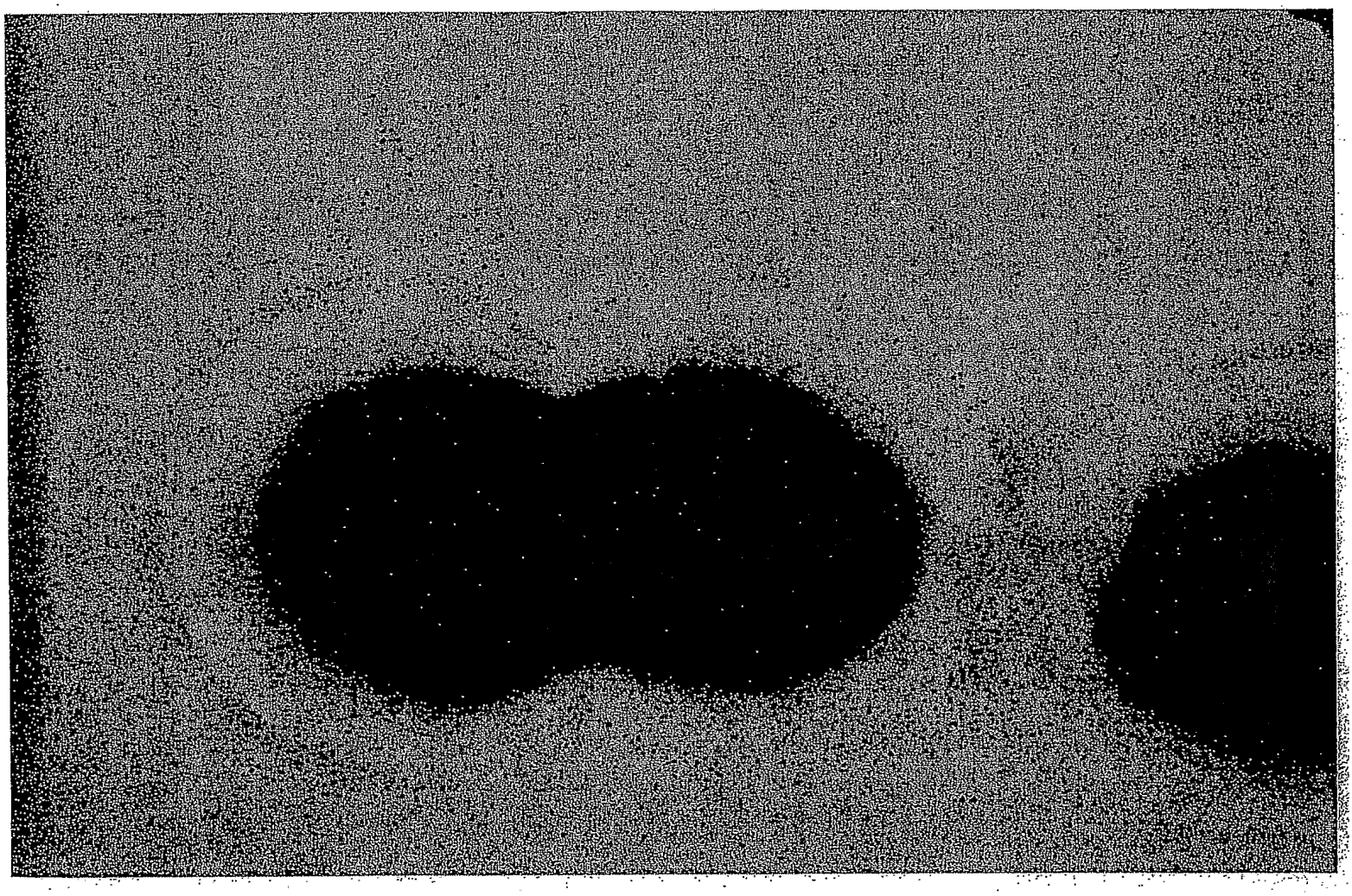


FIGURE 121

313/487

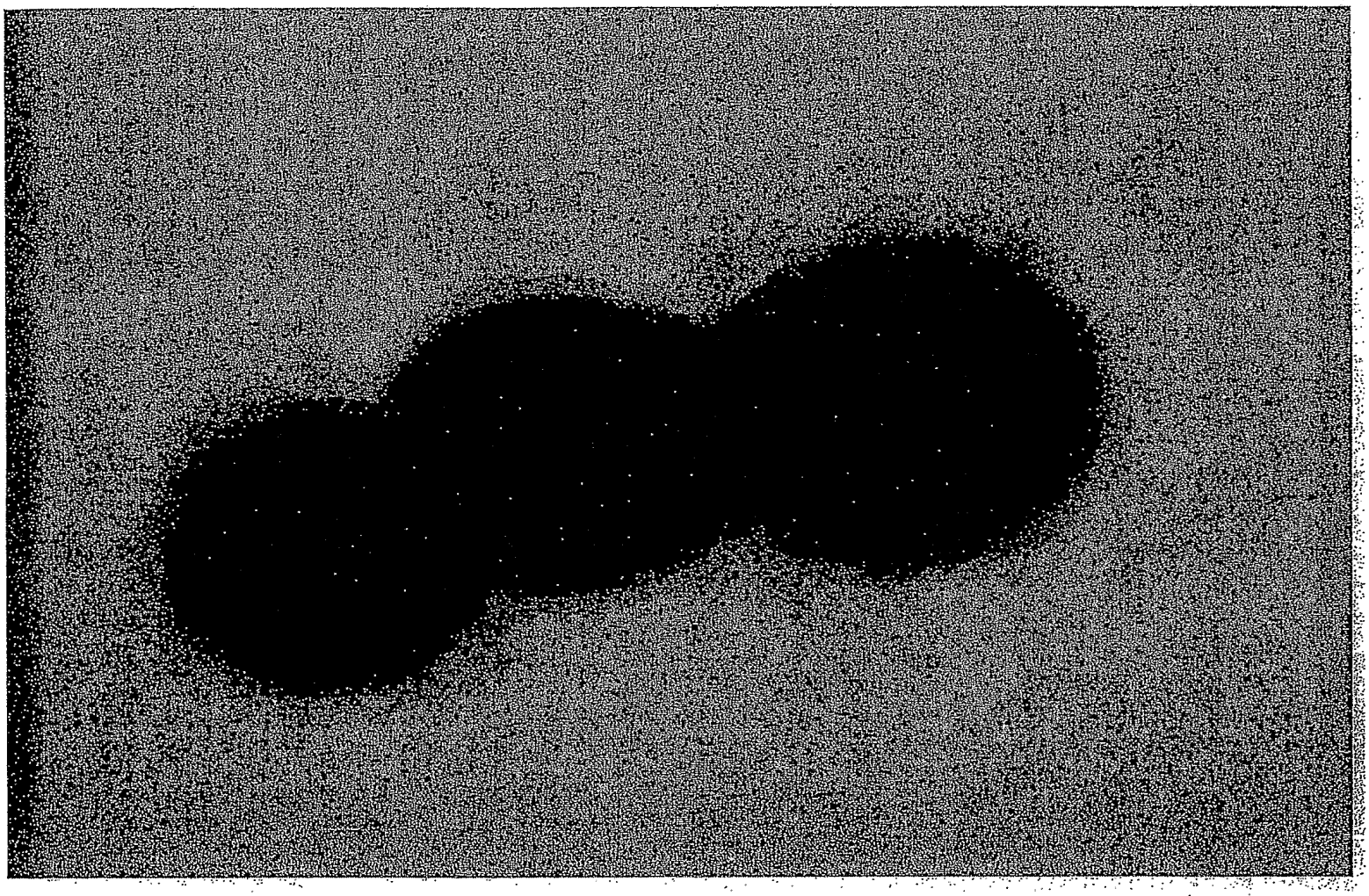


FIGURE 122

PCT/US05/27239

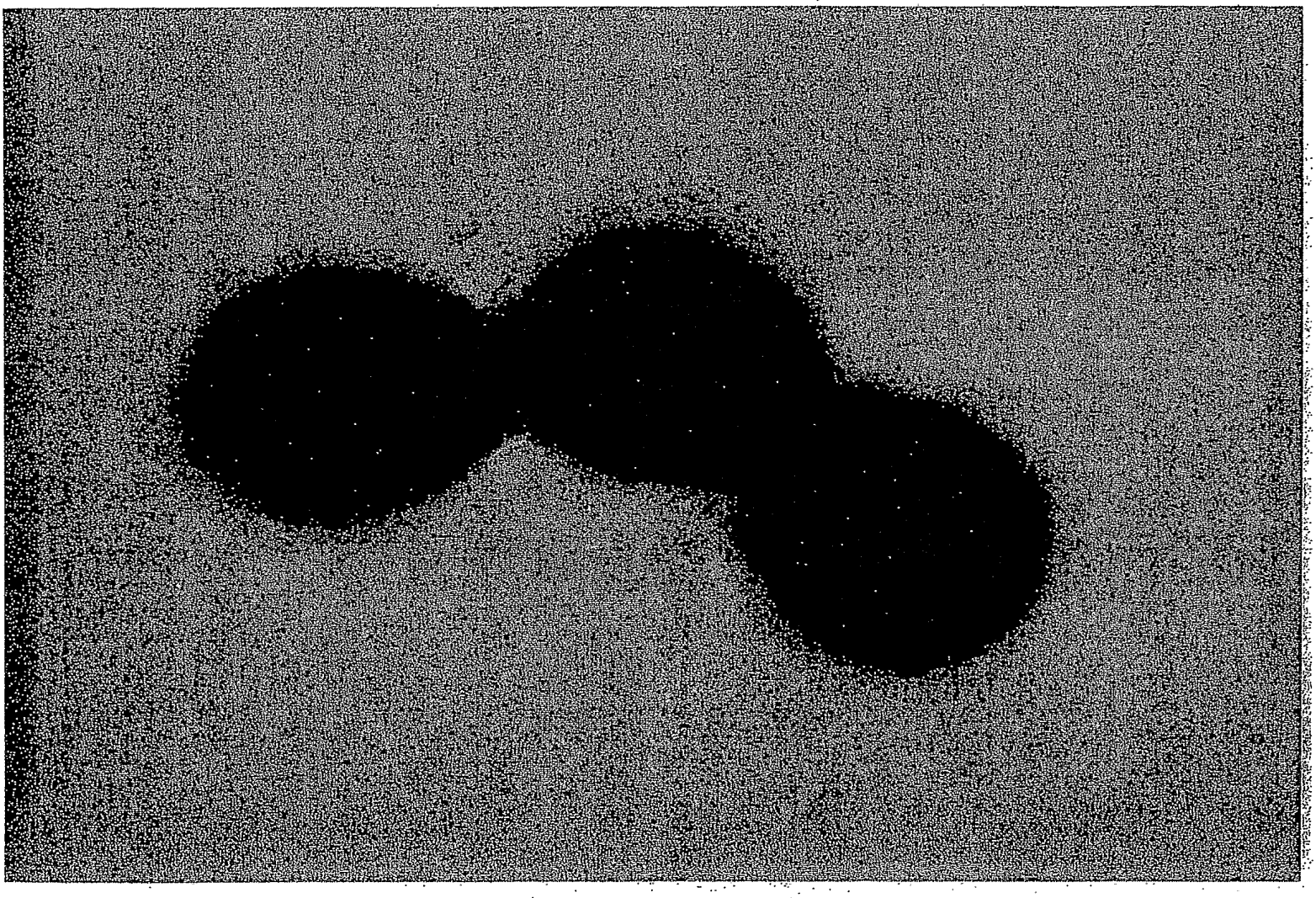


FIGURE 123

PCT/US05/27239

315/487

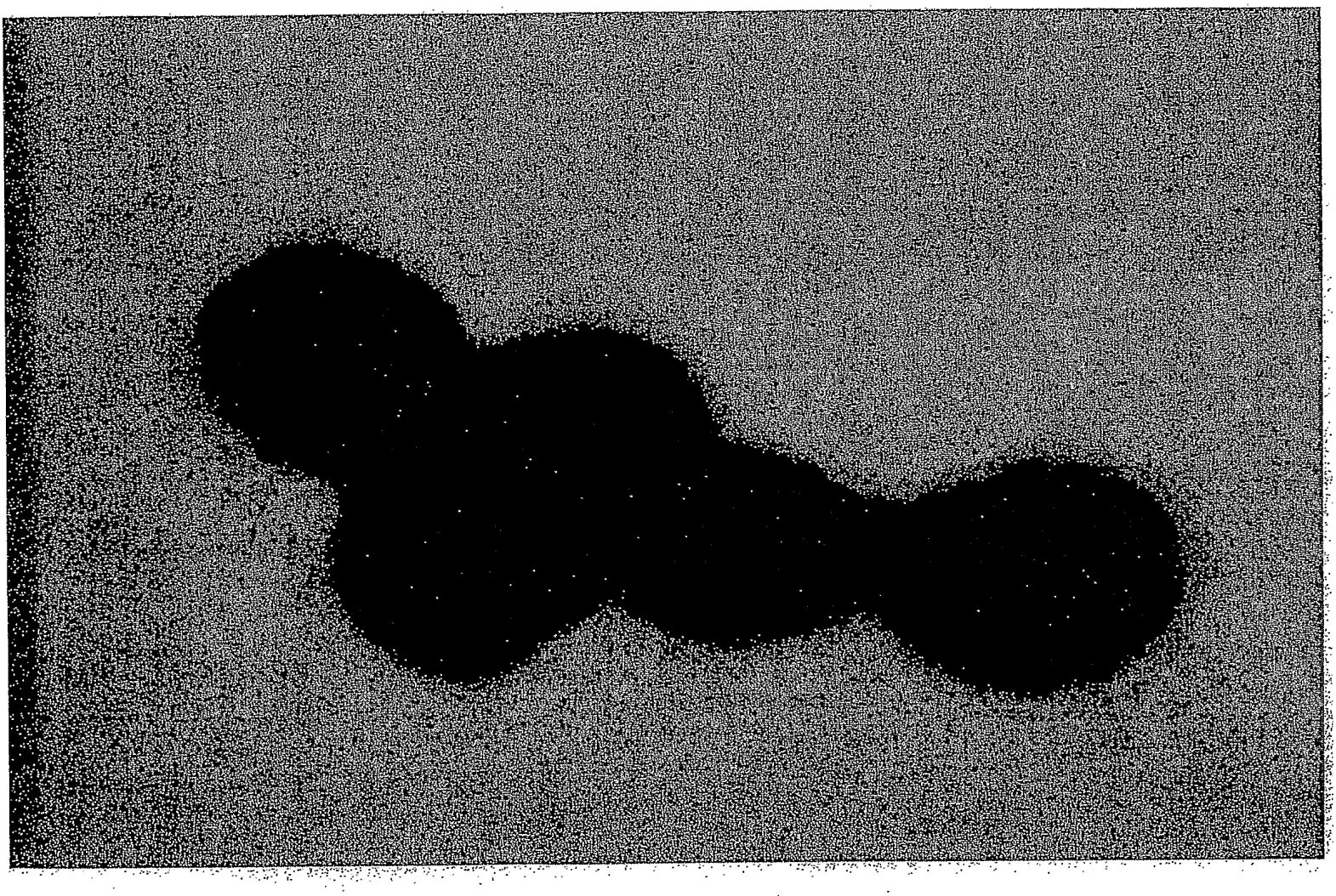


FIGURE 124

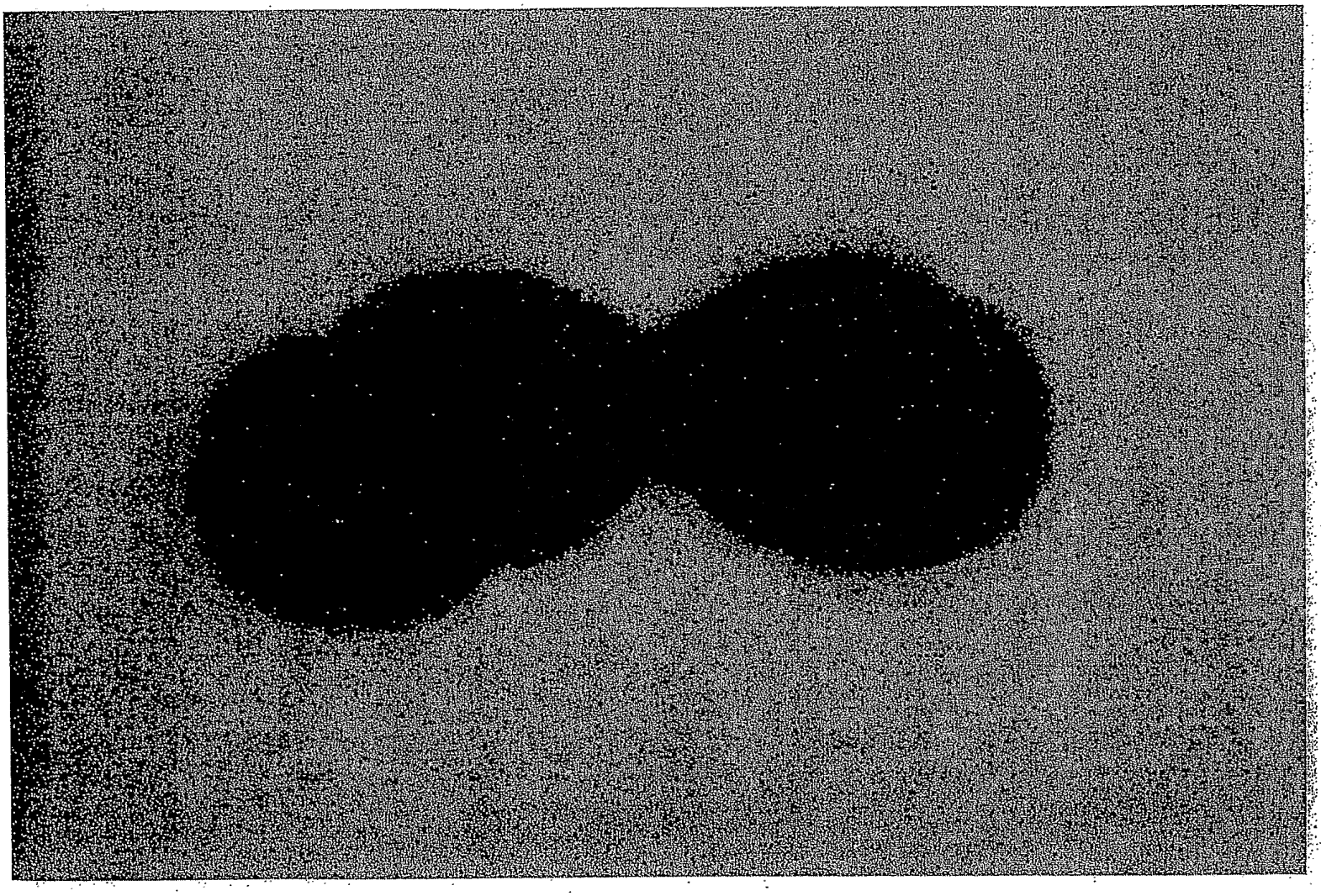


FIGURE 125

PCT/US05/27239 317/487

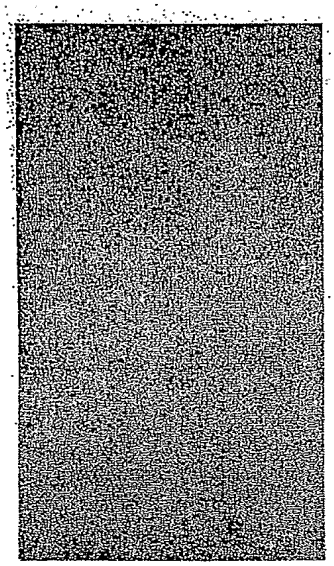


FIGURE 126

Figure 127

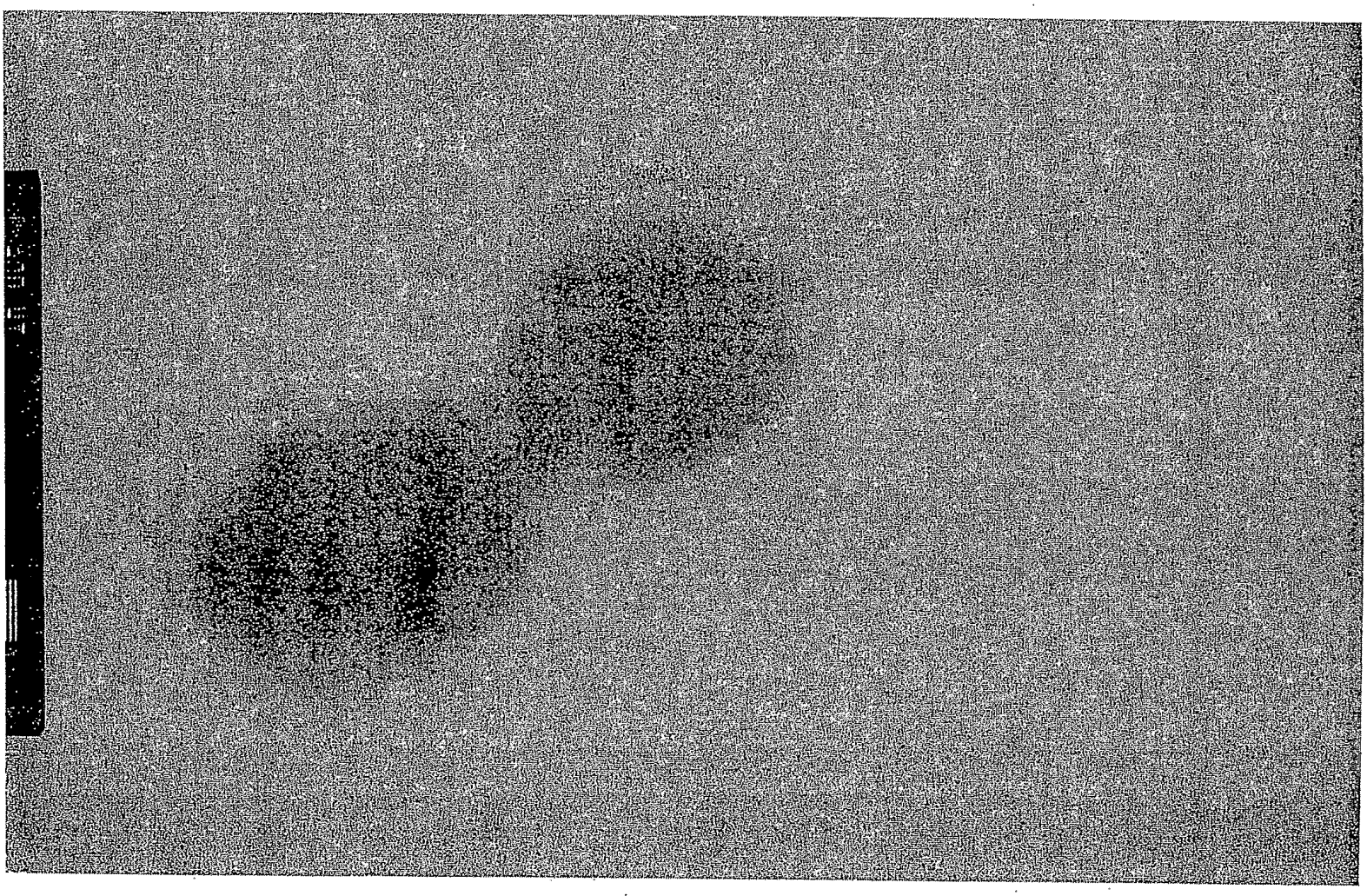


Figure 128

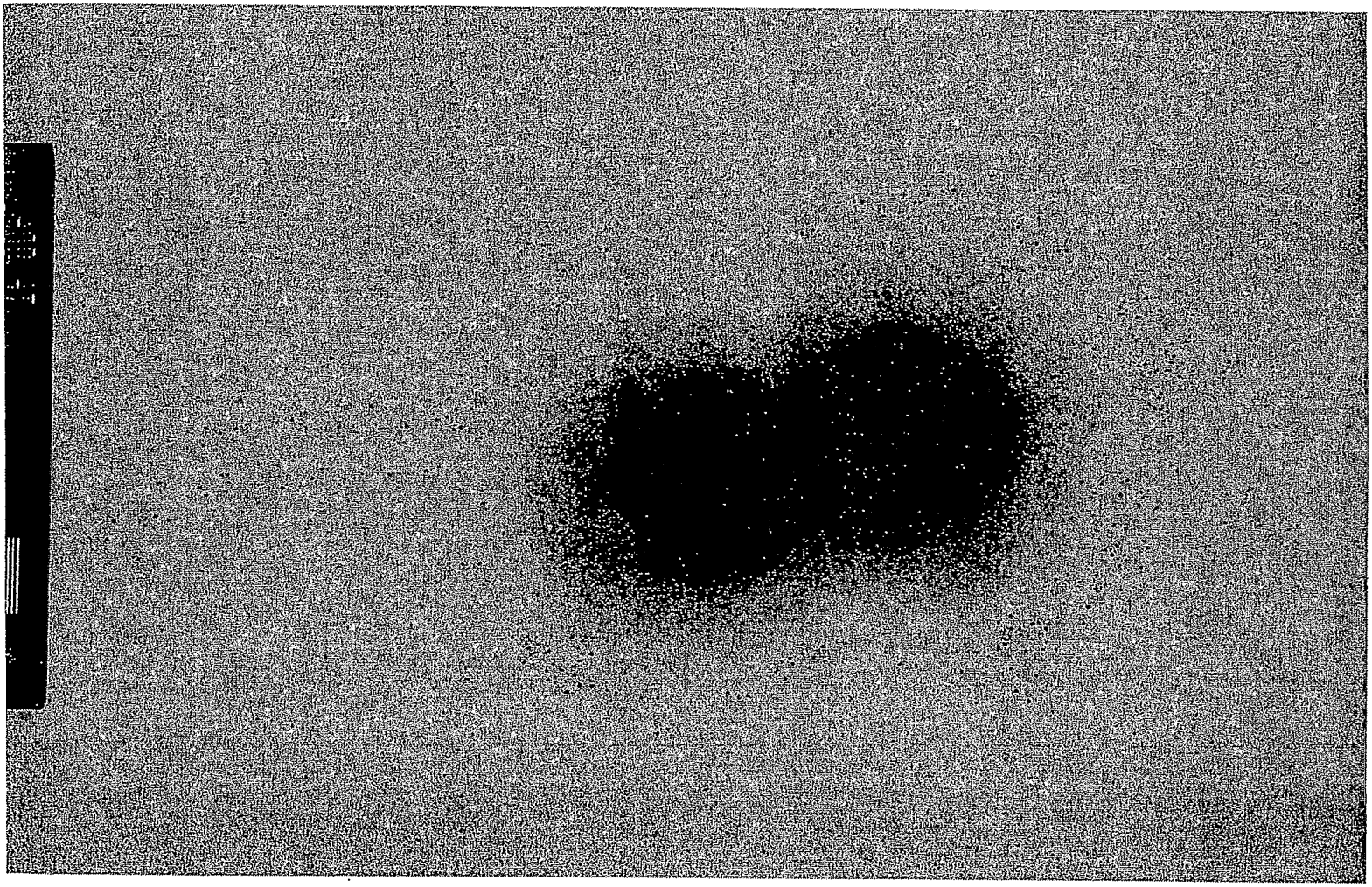


Figure 129